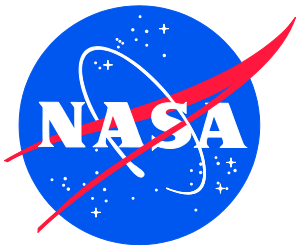
NASA/TM–20220003648
NESC-TI-21-01657



Guidebook for Assessing Similarity and Implementing Empirical Transfer Functions for Probability of Detection (POD) Demonstrations for Signal Based Nondestructive Evaluation (NDE) Methods

Ajay Koshti
Johnson Space Center, Houston, Texas

Peter A. Parker
Langley Research Center, Hampton, Virginia

David S. Forsyth
NDTAnalysis, St John, U.S. Virgin Islands

Michael W. Suits, James L. Walker
Marshall Space Flight Center, Huntsville, Alabama

William H. Prosser/NESC
Langley Research Center, Hampton, Virginia

March 2022

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at *http://www.sti.nasa.gov*

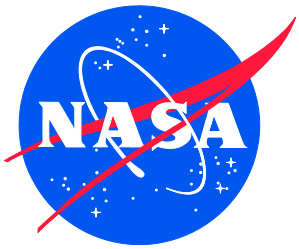
- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:

NASA STI Information Desk
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

NASA/TM–20220003648
NESC-TI-21-01657



Guidebook for Assessing Similarity and Implementing Empirical Transfer Functions for Probability of Detection (POD) Demonstrations for Signal Based Nondestructive Evaluation (NDE) Methods

Ajay Koshti
Johnson Space Center, Houston, Texas

Peter A. Parker
Langley Research Center, Hampton, Virginia

David S. Forsyth
NDTAnalysis, St John, U.S. Virgin Islands

Michael W. Suits, James L. Walker
Marshall Space Flight Center, Huntsville, Alabama

William H. Prosser/NESC
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

March 2022

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

Preface

Purpose – This document provides guidance for assessing similarity relative to nondestructive evaluation (NDE) detectability based on Probability of Detection (POD) demonstrations when comparing different NDE inspection situations. Differences in inspection situations addressed by the similarity assessment methodology herein can include differences in the component being inspected (e.g., different materials, finish, geometry, etc.), and differences in the conditions for the inspection (e.g., different access, temperature, or other environmental conditions, etc.). Additional differences in the inspection situation relative to differences in the method or procedure (e.g., different equipment, sensors, data processing and analysis, etc.) are not addressed in this document but can be assessed using NDE calibration or instrument standardization specimens. When different inspection situations are assessed as similar, existing POD demonstration testing results for one configuration can be applied to the other inspection application and conditions. When the different inspection situations are not similar, it may be possible to extend existing POD results through the use of a transfer function. This document also provides guidance for the use of empirical based transfer functions.

Scope – This document is specifically applicable to NASA programs and projects where NASA Standard and Special NDE methods are used to inspect fracture critical human spaceflight metallic hardware. However, NASA programs and projects that are not human rated may choose to use these methods and the methods may be applicable to other material/flaw types. The guide is restricted to NDE methods that provide signal response data (i.e., not applicable to NDE methods that provide only a hit-miss response) for which the signal is a function of the flaw size.

Audience – This document is written for practicing NDE engineers that are tasked to justify the application of Standard NDE flaw sizes in the inspection of specific spaceflight hardware or are required to design a Special NDE demonstration involving a transfer function. The conceptual approach presented is based on current practice, and therefore, it is expected to be familiar to NDE engineers. This guidebook augments current practice with statistical modeling to provide quantitative metrics that support the NDE engineer's recommendations. Implementing the analysis techniques in this guidebook assumes that the NDE engineer has practical experience in statistical modeling to a level typically provided in an undergraduate introductory applied statistics course that includes linear regression and hypothesis testing. However, this guidebook does not require the level of statistical sophistication expected to apply the methods presented in Military Handbook (MIL-HDBK)-1823A (2009). An NDE engineer who is not comfortable with the analysis techniques presented in this guidebook is encouraged to consult the references provided and/or partner with a statistician for guidance and review of specific applications of transfer function development.

Table of Contents

List of Figures

List of Tables

Nomenclature

AFRL	Air Force Research Laboratory
CIFS	Critical Initial Flaw Size
D	Depth
EDM	Electrical Discharge Machining
FAA	Federal Aviation Administration
FCB	Fracture Control Board
L	Length
LxD	Length Times Depth
MAPOD	Model-Assisted Probability of Detection
MIL-HDBK	Military Handbook
NDE	Nondestructive Evaluation
NDI	Nondestructive Inspection
NDT	Nondestructive Testing
NIST	National Institute of Standards and Technology
PEM	Point Estimate Method
PFIB	Plasma Focused Ion Beam
POD	Probability of Detection
RMSE	Root Mean Squared Error
STD	Standard

Definitions (adapted from NASA Standard (STD) 5009B (2019) - changes and additions specific to Transfer Functions indicated in *italics*)

Blinded Demonstration Inspection – Demonstration inspections performed where the inspector has no knowledge of the flaw location on the specimen.

Capability Demonstration Specimens: A set of specimens made from material similar to the material of the hardware to be inspected with known flaws used to estimate the capability of indication detection, (i.e., Probability of Detection (POD) or other methods of capability assessment) of a nondestructive evaluation (NDE) method.

Cracks or Crack-Like Flaws: A discontinuity assumed to behave like a crack for assessment of material or structural integrity. *Referred to as induced flaws, whether naturally occurring or laboratory simulated.*

Critical Initial Flaw Size (CIFS): The naturally occurring flaw size that is assumed to exist in the part for damage tolerance analysis that is required to be detectable with 90/95 POD.

Defect: One or more flaws whose aggregate size, shape, orientation, location, or properties do not meet specified acceptance criteria and are rejectable.

Flaw: An imperfection or discontinuity that may be detectable by nondestructive testing and is not necessarily rejectable. Examples of flaws in metallic articles include cracks, deep scratches and sharp notches that behave like cracks, material inclusions, forging laps, welding incomplete fusion, penetration, and slag or porosity with a crack-like tail. For additive manufactured metallics, skipped layers, thermal or stress induced cracks, or inclusions are examples.

Hit-Miss NDE Data: Data resulting from an NDE inspection where only the determination of whether an indication is present or not is recorded. Thus, the data at each measurement point corresponds to either a yes or no, or is sometimes represented numerically as a 1 (i.e., indication present) or 0 (no indication). No signal measurements from any NDE sensor output are recorded.

Initial Crack (Flaw) Size: The crack size that is assumed to exist in the part for damage tolerance analysis.

Instrument Calibration: Comparison of an instrument response with, or adjustment of an instrument response to, known references often traceable to the National Institute of Standards and Technology (NIST). This is usually performed periodically, typically at a 1-year interval. After completing calibration, a calibration sticker with calibration expiration date is affixed to the instrument.

Instrument Standardization: Adjustment of an NDE instrument response using an appropriate reference standard with known size discontinuities such as electro-discharged machined slots and flat bottom holes, to obtain or establish a known and reproducible response. This is usually done prior to an examination but can be carried out anytime there is concern about the examination or instrument response. It is also commonly known as calibration prior to initiating an NDE procedure. Instrument standardization should be carried out using a minimum of three data points demonstrating expected correlation between signal response and discontinuity size.

Naturally Occurring Flaw: A flaw that is present in a component as a result of the normally occurring manufacturing processes or usage of the component.

Nondestructive Evaluation (NDE), Nondestructive Inspection (NDI), Nondestructive Testing (NDT): The development and application of technical methods to examine materials or components in ways that do not impair future usefulness and serviceability in order to detect, locate, measure, and evaluate flaws; to assess integrity, properties, and composition; and to measure geometrical characteristics.

NDE Procedure: A written plan providing detailed information on 'how-to' perform a hardware-specific inspection.

*NDE Simulated Fabricated Flaw**:** A flaw that is intentionally placed in a component for the purpose of generating an NDE signal response. These can be produced by a variety of material removal processes (e.g., cutting, drilling, electrical discharge machining (EDM), laser notching, plasma focused ion beam (PFIB) notching, etc.) or additive material forming processes.*

*NDE Simulated Induced Flaw**:** A flaw that is intentionally placed in a component for the purpose of generating an NDE signal response. Induced flaws are produced by intentional loading (thermal, mechanical, etc.) to induce damage (e.g., cracks, delaminations, disbonds, etc.).*

NDE Transfer Function: A function that describes the relationship between signal responses for an NDE method as a function of flaw size for different types of flaws (e.g., naturally occurring flaws, load induced or material removal NDE simulated flaws) or for flaws in different types of components (e.g., simple geometries such as cylinders or flat plates or structural component of interest with complex geometry).

Signal Response NDE Data: Data from an inspection where the NDE sensor produces a signal output (e.g., voltage, current, etc.) that is measured and proportional to flaw size. The determination for whether an indication is present is typically made based on a threshold value of the signal response NDE data.

Special NDE: Nondestructive inspections of fracture-critical hardware that are capable of detecting cracks or crack-like flaws smaller than those assumed detectable by Standard NDE or do not conform to the requirements for Standard NDE as set forth in NASA Standard 5009B. Special NDE methods are not limited to fluorescent penetrant, radiography, ultrasonic, eddy current, and magnetic particle.

Standard NDE: NDE methods of metallic materials for which a statistically based flaw detection capability has been established. Standard NDE methods addressed by NASA Standard 5009B are limited to the fluorescent penetrant, radiographic, ultrasonic, eddy current, and magnetic particle methods employing techniques with established capabilities.

Similarity: The outcome of an assessment that the same POD is expected in different NDE inspection situations that might include variations in NDE method/procedure, components being inspected, and/or inspection conditions.

Target Flaw Size: Flaw size that is established for which the NDE method and inspector must provide reliable (i.e., 90/95 POD) detection. This flaw size is then used in POD demonstration testing.

Unblinded Demonstration Inspection – Demonstration inspections performed where the inspector has knowledge of the flaw location on the specimen.

1.0 Introduction

NASA STD 5009B specifies that nondestructive evaluation (NDE) methods/inspectors used to inspect fracture critical metallic components in human-rated aerospace flight systems provide 90-percent probability of detection (POD) with 95 percent confidence (i.e., 90/95 POD). This can be accomplished in two ways. The first is to use a designated Standard NDE method and the established Standard NDE flaw sizes contained within NASA STD 5009B along with the prescribed industry standard technique processes. For all other situations, the NDE method/inspector is designated as Special NDE and currently the 90/95 POD must be demonstrated by POD testing. Since NASA STD 5009B is limited to metallic components, the methodology presented in this guidebook may not apply to nonmetallic or composite materials.

The NASA Standard NDE flaw sizes were established by an extensive POD demonstration testing program, see Bishop (1973), decades ago performed for the Space Shuttle Program orbiter fracture control development. These tests were performed on one material alloy using flat panels with fatigue cracks, and a limited range of thicknesses. To extend these results and avoid time-consuming additional POD demonstration testing unless necessary, NASA STD 5009B discusses (but does not define) the concept of similarity or transferability. It notes that when operational inspection conditions deviate significantly from the conditions of the original Standard NDE POD testing, the transferability or similarity of the resulting documented detectable flaw sizes should be evaluated and verified by documented evidence (e.g., experimental data or other available test data documentation). Thus, if the differing inspection conditions are judged ‘similar’, then the demonstrated detectable flaw sizes catalogued in NASA STD 5009B can be ‘transferred’ to the real inspection application. While NASA STD 5009B only discusses similarity relative to the component being inspected, there are many additional factors that can affect the detectability of an NDE inspection. These include the NDE method and procedure, the equipment used, and the conditions under which the inspection is performed (e.g., environmental and access conditions). Thus, a similarity assessment needs to evaluate the effects of differences in method/procedure, component being inspected, and inspection conditions. Although only discussed relative to Standard NDE, it is inferred from NASA STD 5009B that if two Special NDE inspection conditions are similar, then POD demonstration test results from one would be applicable to the other.

While NASA STD 5009B describes the use of documented evidence (e.g., experimental data to assess similarity), no formal process or procedure is provided for conducting a similarity assessment. Further, historically, engineering judgement has often been the basis for declaring similarity with little, if any, supporting experimental data. This document provides guidance for the assessment of similarity relative to variations in the component being inspected and the conditions of the inspection. The use of calibration or instrumentation standardization specimens can be used for the assessment of similarity with respect to variations in the NDE method and procedure.

When it is assessed that a different NDE inspection condition is not similar, it may be possible to use the established POD demonstration testing results through the use of an NDE transfer function. An NDE transfer function describes the relationship between signal responses for an NDE method as a function of flaw size for different types of flaws (e.g., fatigue cracks versus electrical discharge machining (EDM) notches) or similar flaws in different types of components (e.g., flat plates versus a structural component of interest). While NASA STD 5009B only discusses transferability

when similarity is established, transfer functions for non-similar inspection conditions have been used within NASA on a limited case-by-case basis with approval from the responsible Fracture Control Board (FCB). This guidebook is intended for use by practicing NDE engineers identified in the Preface to provide guidance for a transfer function methodology that can be used, and it is not intended to exclude other approaches for consideration and possible continued use.

2.0 Literature Review

Transfer functions are an element of model-assisted probability of detection (MAPOD) evaluation, which is an emerging research and development area within the NDE discipline. The MAPOD Working Group was established in 2004 by the Air Force Research Laboratory (AFRL) in cooperation with the Federal Aviation Administration (FAA) and NASA, see Thompson et al. (2009). The goal of the working group was to promote increased understanding and adoption of MAPOD. Their charter spanned experimentally assessed and physics-based modeling approaches. MIL-HDBK-1823A incorporated the contributions of the MAPOD Working Group in Appendix H, and it provides an introduction and conceptual overview of designing and analyzing a transfer function study. However, it lacks the specificity necessary for many NDE engineers to design, execute, and analyze a transfer study, reflecting the state of the research area at the time of publication. Meyer et al. (2014) offers a comprehensive review of literature on MAPOD, and the authors categorize empirical transfer functions as those based on experimental inspection measurements and full model-assisted transfer functions that incorporate physics-based NDE method modeling. This reference primarily focuses on analysis approaches, many of which are statistically advanced. Lindgren et al. (2018) includes a section devoted to MAPOD and highlights the importance of integrating experimental and physics-based approaches. The authors suggest that empirical approaches may be limited to minor differences in material, flaw type, or shape changes, and that physics-based modeling offers the ability to characterize more significant transfer factors. This article includes two condensed case studies of transfer functions used on aircraft engine components and structures. It suggests that while a sufficient sample size of naturally occurring flaws in flight components is rarely available to perform a POD study, a few flight service components can serve as a vital validation check of a transfer function.

In terms of methodology and case studies in the literature, Koshti (2021a) proposes an approach to transfer the 90/95 flaw size from different flaw and specimen type configurations. The method utilizes a MIL-HDBK-1823A $\hat{a}$-versus-a POD study approach for various flaw and specimen cases with reduced sample sizes for efficiency compared to MIL-HDBK-1823A's recommendations. It proposes a method to propagate the uncertainty from inspected specimen configurations to predict the 90/95 flaw size for a flight component. Jones (2015) et al. discusses the evaluation of an eddy current probe under challenging inspection site access constraints for an aircraft bulkhead where the inspector cannot see the probe or inspection site and is required to be in a prone position during the inspection. The authors add an experience-based margin of conservatism to the transferred signal versus flaw size relationship for inspections. Bode et al. (2008) and (2012) report on a transfer function developed for ultrasonic inspection of riveted lap joints in an airframe. This work includes naturally occurring flaws in a flight component from a retired aircraft to validate the transfer function. Harding et al. (2007), (2008), and (2009) present a transfer function study on flaws at fastener holes in aircraft wings, and they develop an empirical transfer function for an ultrasonic inspection method. This work is the best illustration found in the current literature of an empirical transfer function development.

In summary, the literature does not offer a general approach to develop an experimentally assessed transfer function, and it is dominated by physics-based modeling transfer functions to estimate a full POD relationship over a range of flaw sizes, typically related to transferring flaw types. When transfer functions based on experimental inspection measurements are presented, there are no examples in the literature where the modeling approach first considers if a transfer function can be reliably detected and estimated in the presence of inspection variability. Seeking to address these gaps in the existing literature, this guidebook proposes a general methodology for experimentally derived transfer functions across specimen types that quantitatively assesses if a transfer function is justified or whether similarity (i.e., a 1-to-1 transfer function) is supported. The approach presented in this guidebook could be considered a point transfer function since it transfers a flaw size across specimen configurations for Special NDE demonstration required by 5009B, but it does not transfer the full POD relationship.

3.0 Flaw and Specimen Types

For discussions of similarity, transfer functions and POD testing, it is necessary to distinguish between the different types of flaws that might be involved. Naturally occurring flaws originate during the production or usage of the component. These are what NDE engineers seek to detect when performing an NDE inspection during the manufacturing, assembly, and testing processes or periodically after usage of the component. However, seldom, if ever, is the NDE engineer able to gather a set of the needed numbers and sizes of naturally occurring flaws, that are independently characterized by other NDE methods, such that the NDE engineer can use them for POD demonstration testing.

Thus, NDE simulated flaws are required. These simulated flaws can be induced by applying mechanical or thermal loads to the component or fabricated through the use of additive or mechanical removal techniques. Ideally, NDE simulated induced flaws provide signal responses of similar magnitude and variability to that which would be expected from similar sized naturally occurring flaws. Historically, cracks produced by intentional loading (e.g., laboratory produced fatigue cracks) have been used as simulated induced NDE flaws in metals when the naturally occurring flaw desired to be detected is a crack or crack-like flaw as they provide several advantages for POD demonstration testing. They have natural variability in crack morphology that produces NDE signals with variability that may be representative to that from naturally occurring flaws. In addition, cracks produced by methods (e.g., fatigue) may be tightly closed and produce signals that are of similar or smaller magnitude to naturally occurring flaws. However, it is noted that different methods of loading can produce different degrees of crack opening for the same ‘size’ (i.e., length, depth, surface area, and morphology) cracks. This variation in opening may have an influence on the NDE signal that results and thus the detectability of the NDE simulated cracks. It is recommended that when such cracks are used as induced flaws, the crack opening dimension is characterized, and it is evaluated to ensure that it is consistent, or smaller than that of the naturally occurring crack-like flaw to be detected.

In many situations, however, particularly in a flight component with complex geometry, induced flaws cannot be produced, or at least not in a controlled manner to facilitate the number of flawed specimens required for POD testing. In such cases, fabricated flaws produced by material removal methods (e.g., cutting, drilling, EDM, laser notching, plasma focused ion beam (PFIB) notching, etc.) or by additive manufacturing methods may be produced in the component of interest. These types of flaws have the advantage of being relatively easy to produce with controlled dimensions

even in complex geometry components. However, they have disadvantages when used for POD demonstration testing. They typically produce signals that are much larger in amplitude than a similar size naturally occurring or induced NDE simulated flaw. Additionally, they may not provide as much variability in morphology and thus do not provide as much variability in signal response as naturally occurring flaws.

To address the limitations of using fabricated flaws for POD testing, NDE transfer functions can be developed from a comparison of signal responses from flaws produced by material removal and loading methods in specimens with simple geometry (e.g., cylinders or flat plates), or from a comparison of signal responses from a given flaw type in different types of specimens (e.g., simple specimens including flat plates versus flight components). These NDE transfer functions can be used to relate signal responses between different types of flaws and flaws in different component geometries.

Focusing on differences in specimen configurations, Figure 3.0-1 shows the four possible combinations of flaw type and specimen type. The upper left combination represents naturally occurring flaws or simulated induced flaws (assumed to be conservative to the naturally occurring flaws) in flight components. If these specimens are available, then direct POD demonstration testing can be performed, and a transfer function approach is not required. However, this combination is not always feasible in practice, and therefore an evaluation of similarity and the development of a transfer function is often required. Two types of transfer functions are illustrated. One is a transfer function across flaw types, where data from simulated induced and simulated fabricated flaws are obtained and evaluated to develop the transfer function, which is represented by the bottom row of the diagram. This transfer function is typically used to size simulated fabricated flaws in the component type specimen to perform a POD demonstration. The underlying assumption is then that the transfer function is independent of the type of specimen. This transfer function has been used on a limited basis with FCB approval within NASA. However, detailed guidance on how to implement flaw type transfer functions for NASA fracture critical applications has not been documented. Additionally, one concern with this approach is that the POD demonstration testing is performed with simulated fabricated flaws, which may not provide representative variability in signal response as naturally occurring or induced flaws. Therefore, the resulting transfer function and demonstrated POD flaw sizes may not be conservative. As such, an alternate transfer function methodology is recommended in this guidebook. However, given that flaw type transfer functions have been used for NASA programs and may be used in the future, specific guidance for flaw type transfer functions that incorporates steps to better capture the variability due to induced flaws is included in Appendix D.

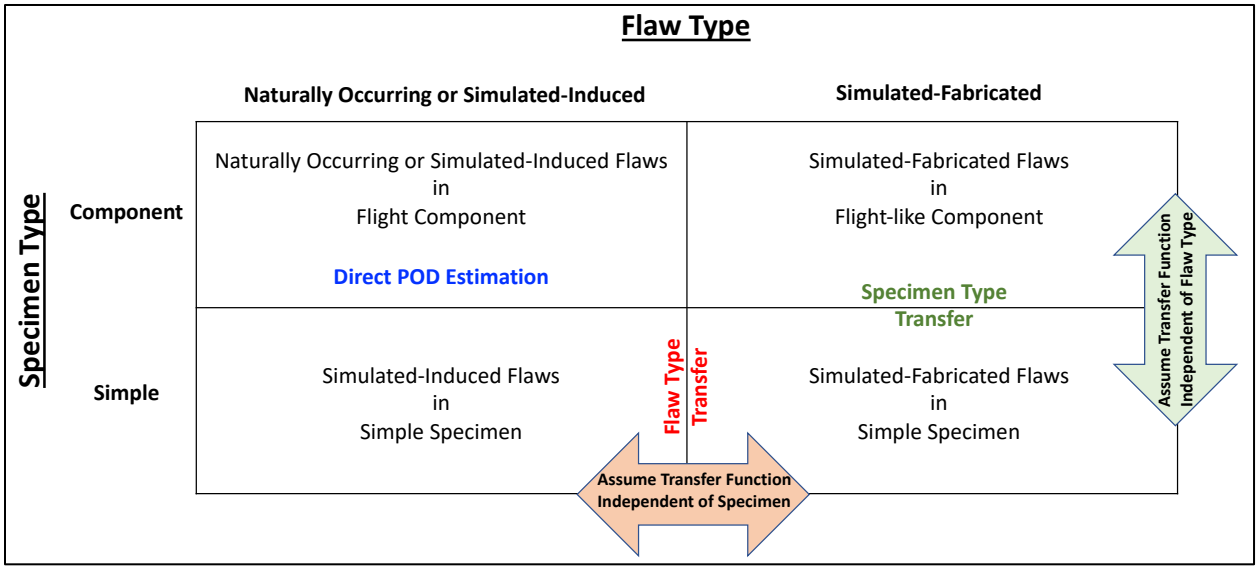


Figure 3.0-1. Block diagram of relationships between flaw and specimen types and two types of possible transfer functions and underlying assumption for each.

Due to the concerns about the lack of conservatism with the flaw type transfer function approach, an alternative approach is based on a specimen type transfer function. In this case, signal data are obtained from the same type of simulated fabricated flaws in the simple and component specimen configurations. In this case, the transfer function is then assumed to be independent of flaw type. This specimen type transfer function can be used in two ways. When there is an established detectable flaw size from POD testing in one inspection configuration (e.g., flat plate specimens), the transfer function can be used to calculate the estimated detectable flaw size in another inspection configuration (e.g., structural component of interest). This avoids additional POD demonstration testing, which can add schedule delays to a program or project. The transfer function can also be used when the NDE engineer has an identified critical initial flaw size (CIFS) for a flight component of interest but cannot produce the necessary simulated flaws in that component to perform the required NDE POD demonstration testing. In this case, the transfer function is used to size the flaws for simplified demonstration specimens to perform POD testing. In either case, the underlying POD demonstration testing is performed with simulated induced flaws that are assumed to be conservative to the naturally occurring flaws of interest with respect to signal response and detectability. The variability in the signal response of simulated induced flaws is expected to be more representative of naturally occurring flaws compared to simulated fabricated flaws.

In this guidebook, the evaluation of specimen similarity and the development of transfer functions is based on comparing NDE inspection measurements using the same technique and inspector from simulated-fabricated flaws in two types of specimens, generically referred to as *Component* and *Simple* specimens. In general, Component specimens are flight articles, or representative of them with respect to geometry, material, surface finish, and relevant operational inspection conditions (e.g., access and environmental conditions). Simple specimens are typically laboratory inspection specimens that may share some attributes of the Component specimens but are not fully representative of a particular flight article. Simple specimens may include flat plates as an approximation for acreage skin material with gentle curvature, cylinders or tubing for pressure vessels or piping, or round bars as a stand-in for struts or fasteners. When possible, it is desirable

for the Simple specimens to share as many attributes (e.g., material type, surface finish, etc.) with the Component specimen as possible. However, issues (e.g., material cost, availability, machinability, etc.) may limit the commonality that can be achieved.

4.0 Application of Transfer Functions

When similarity cannot be established, it may be possible to utilize established POD flaw sizes through an adjustment to the 90/95 flaw size based on a transfer function. Such a transfer function can be used to define a target flaw size for POD testing in a Simple specimen from a desired CIFS in a Component. These two different uses of a transfer function are designated the *Forward* or *Inverse* application case. In the Forward case, a naturally occurring CIFS in a flight component is specified, which is often derived from the CIFS determined by fracture analysis or test, and a representative simulated-induced flaw size is to be specified in a Simple specimen for Special NDE POD demonstration by each inspector to be qualified. Figure 4.0-1 presents a block diagram of the Forward case. As a Forward example used throughout the presentation of the methods, consider a case where fracture analysis specifies the detection of a naturally occurring flaw size of 0.065 in the flight component. What is the simulated-induced flaw size in a Simple specimen that would be used in a Special NDE POD demonstration test? In the examples that follow, the flaw size units are non-dimensional.

In the Inverse application case, an established simulated-induced flaw size with quantified POD, often 90/95, from a Simple specimen is available as the baseline, and it desired to transfer it to an estimated naturally occurring flaw size in a flight component. Figure 4.0-2 presents a block diagram of the Inverse case. As an example of the Inverse case, consider a simulated-induced flaw size of 0.040 in a Simple specimen has been demonstrated to have 90/95 POD. What is the estimated naturally occurring flaw size that could be detected in a flight component?

For the Forward and Inverse cases, transfer function evaluation measurements of simulated-fabricated flaws are used to assess similarity, and are used to estimate the transfer function relationships between Simple and Component specimens. As previously stated, a fundamental assumption in applying a transfer function is that the specimen type transfer applies to all flaw types. In other words, it is assumed that the specimen type transfer does not depend on flaw type, which in statistical terminology assumes that there is no interaction between flaw type and specimen type.

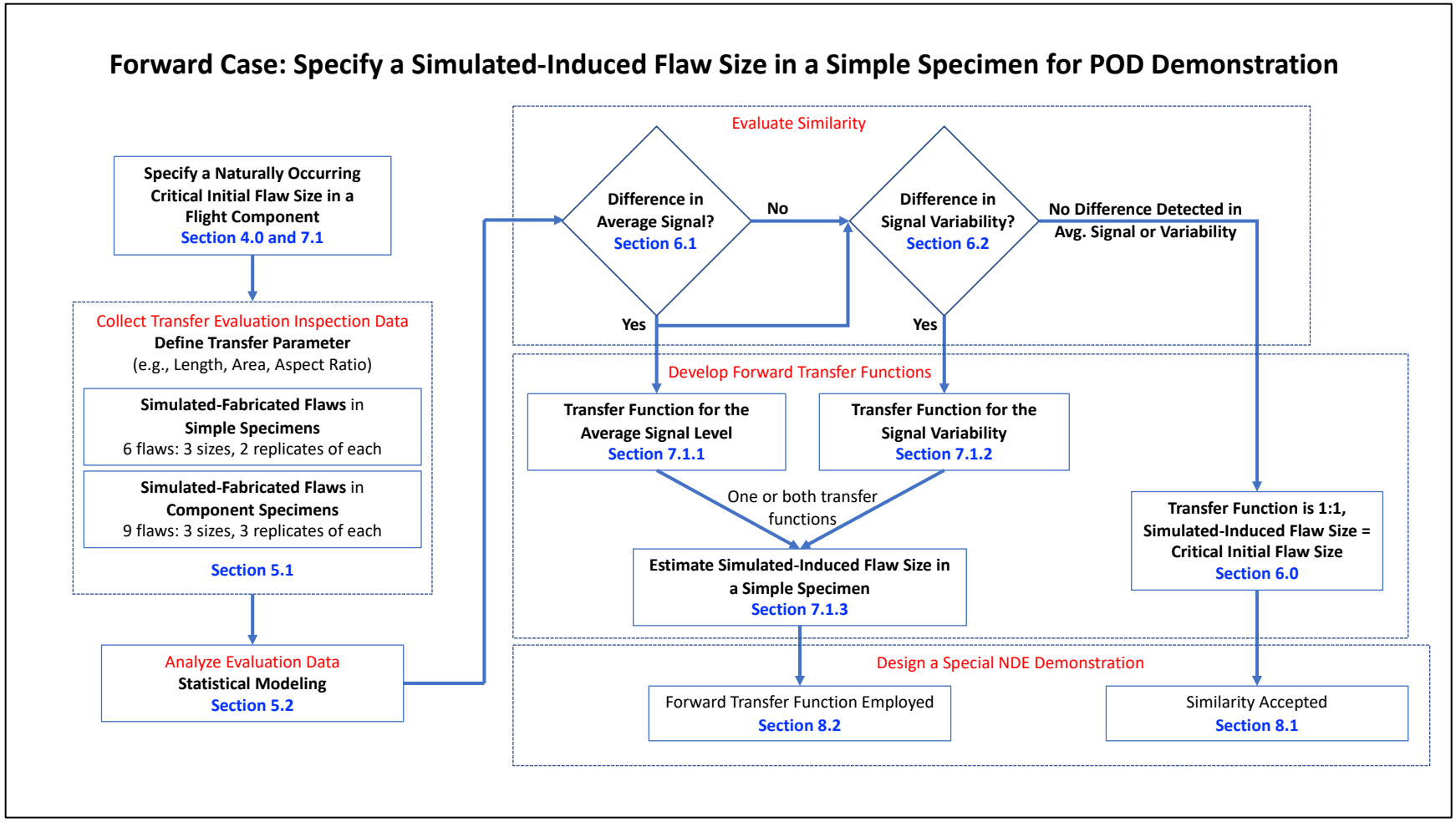


Figure 4.0-1. Forward case block diagram.

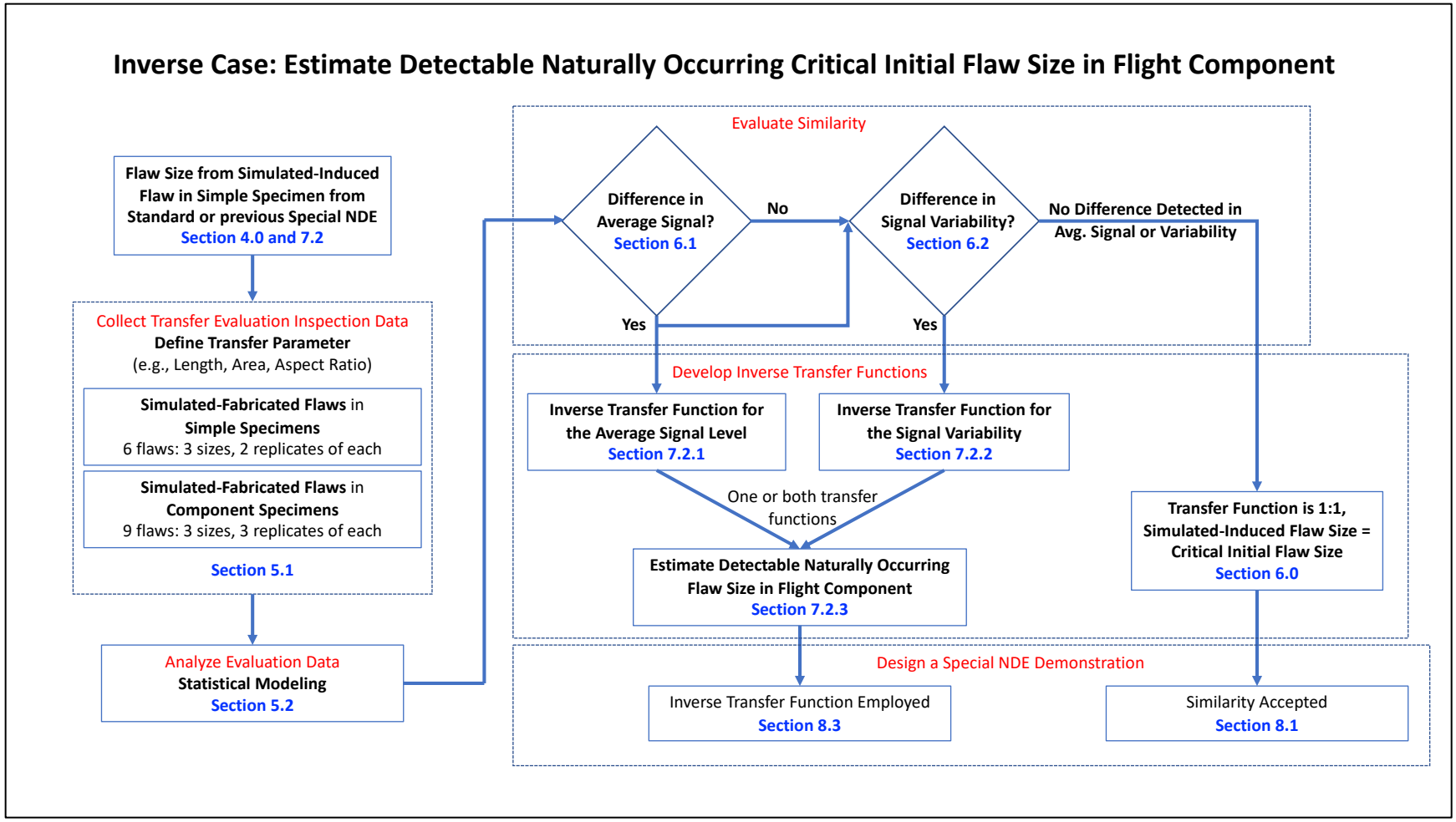


Figure 4.0-2. Inverse case block diagram.

5.0 Assessment of Similarity with Respect to Variations in Inspected Components and Inspection Conditions

Guidance is provided for assessing similarity of different NDE inspection situations with respect to variations in the component being inspected (e.g., differences in material, geometry, condition, etc.) and variations in the inspection conditions (e.g., differences in access and environment under which the inspection is performed). The basis of the method is to develop a transfer function using simulated fabricated flaws of the same sizes in the two different inspection conditions. If within assessed variability of the signal responses it is determined that a transfer function is not required (i.e., the transfer function is 1-to-1), then the different inspection situations are judged to be similar.

5.1 Transfer Function Evaluation Measurements used to Assess Similarity

Transfer function evaluation measurement datasets are used to evaluate similarity and to develop a transfer function. In this methodology, it is assumed that a linear relationship of signal as a function of flaw size can be modeled using an appropriate parameterization of the flaw size (e.g., length, depth, area) and/or an appropriate transformation (e.g., log) of the flaw size and/or the signal. MIL-HDBK-1823A provides guidance on determining appropriate transformations. The general approach outlined in this methodology can be extended to more complex relationships between the signal and flaw size (e.g., higher order polynomial models and multifactor models) that relate signal to two or more factors (e.g., length and depth). Regardless of the model form, extrapolation beyond the range of flaw sizes in which the model is estimated is not recommended.

The inspection signal is denoted as y, and the explanatory variable is the flaw size denoted as x. As presented in this guidebook, a single explanatory variable of flaw size is illustrated, but the principles could be extended to multiple explanatory variables. One inspector is assumed to obtain the transfer function evaluation measurements from the Simple and Component specimens. Therefore, inspector-to-inspector variability is not considered in evaluating similarity or a transfer function. Inspector variability is an important component of POD evaluation, and it is included in the subsequent Special NDE demonstration.

Subscripts on y and x variables denote the combination of flaw type and specimen type, where N indicates a simulated-fabricated flaw such as a notch, and C and S represent a Component or Simple specimen, respectively. In the transfer function evaluation measurement data, an inspection signal from a simulated-fabricated flaw in a Component specimen is y_{NC} and an inspection signal from a simulated-fabricated flaw in a Simple specimen is y_{NS}. Similarly for flaw sizes, x_{NC} and x_{NS} represent simulated-fabricated flaw sizes in Component and Simple specimens, respectively.

Under the linear model assumption, a minimum of three simulated-fabricated flaw sizes distributed over the range of interest are recommended for the transfer function evaluation measurement data in the Simple and Component specimens. It is recommended to use the same nominal flaw sizes in both specimen types. Three unique, nominal flaw sizes provide one additional degree of freedom to assess the fit of the linear model. To estimate the variability in the inspection signal, a minimum of two replicates are recommended at each of the three unique flaw sizes. Replicated flaw measurements are performed on multiple flaws fabricated of the same nominal size that may be spatially distributed on a specimen or placed in multiple specimens. Replicated flaw measurements are not repeated measurements of the same flaw. Using replicated flaws allows for signal variability to be estimated independently from the residuals of the model fit, since the variability

of multiple flaws of the same nominal size is independent of the fitted model. It is assumed that the Component specimen and access restrictions may exhibit higher variability than the Simple specimen, and therefore, three replicates of each flaw size are recommended for the Component specimen.

The flaw sizes used in the transfer function evaluation measurements are specified by the NDE engineer and are informed by prior knowledge of the inspection conditions. As a minimum criteria, the range of flaw sizes used in the transfer function evaluation measurement datasets should cover the CIFS of interest to avoid extrapolation of the models. If historical data or literature examples exist, the NDE engineer may rely on these sources to specify the flaw sizes. In the absence of any prior information, a 1-to-1 transfer function could be assumed between the specimen types. If a 1-to-1 transfer function is assumed, then the average flaw size would be equal to the CIFS that is to be detected in the flight component. To estimate a relationship between signal and flaw size, a range of flaw sizes is required. If historical data or literature examples exist, then these sources can be used to specify a reasonable range of flaw sizes that are above the noise threshold and below the saturation level of the technique. These determinations require expertise in the specific NDE technique and its calibration, and therefore they cannot be provided in general in this guidebook. In the absence of any prior information regarding the range of flaw sizes, the high and low levels of flaw size could be specified as plus and minus one-half of CIFS. As an example illustration, if the CIFS is 0.060, then the flaw sizes used in the transfer function evaluation measurement dataset could be 0.030, 0.060, 0.090 in the Simple and Component specmens. It is anticipated iteration may be required to specify ranges of flaws sizes in both specimen types in order to avoid extrapolating in the transfer function development.

To explain the methodology in this guidebook, simulated transfer function evaluation measurement datasets are provided in Tables 5.1-1 and 5.1-2 based on the minimum recommended sample sizes obtained by one inspector for both specimen types. Figure 5.1-1 is a plot of these measurements with estimated linear models shown by dashed lines. It is best practice and strongly encouraged to plot the raw inspection measurements to identify anomalous measurements and visually confirm that the signal increases with flaw size over the full range of inspections. These are simulated datasets for the purpose of illustrating this methodology, and the replicated flaw sizes are shown as identical to easily identify the experimental structure of the design; however, some variation in fabricated flaw sizes is expected.

Table 5.1-1. Transfer Function Evaluation Data from Simulated-Fabricated Flaws in Simple Specimen

Flaw Size (x_{NS})	Signal (y_{NS})
0.020	13.5
0.020	14.4
0.050	28.2
0.050	29.0
0.080	42.8
0.080	43.2

Table 5.1-2. Transfer Function Evaluation Data from Simulated-Fabricated Flaws in Component Specimen

Flaw Size (x_{NC})	Signal (y_{NC})
0.020	8.8
0.020	15.1
0.020	7.0
0.050	22.5
0.050	21.7
0.050	14.5
0.080	31.1
0.080	37.8
0.080	25.4

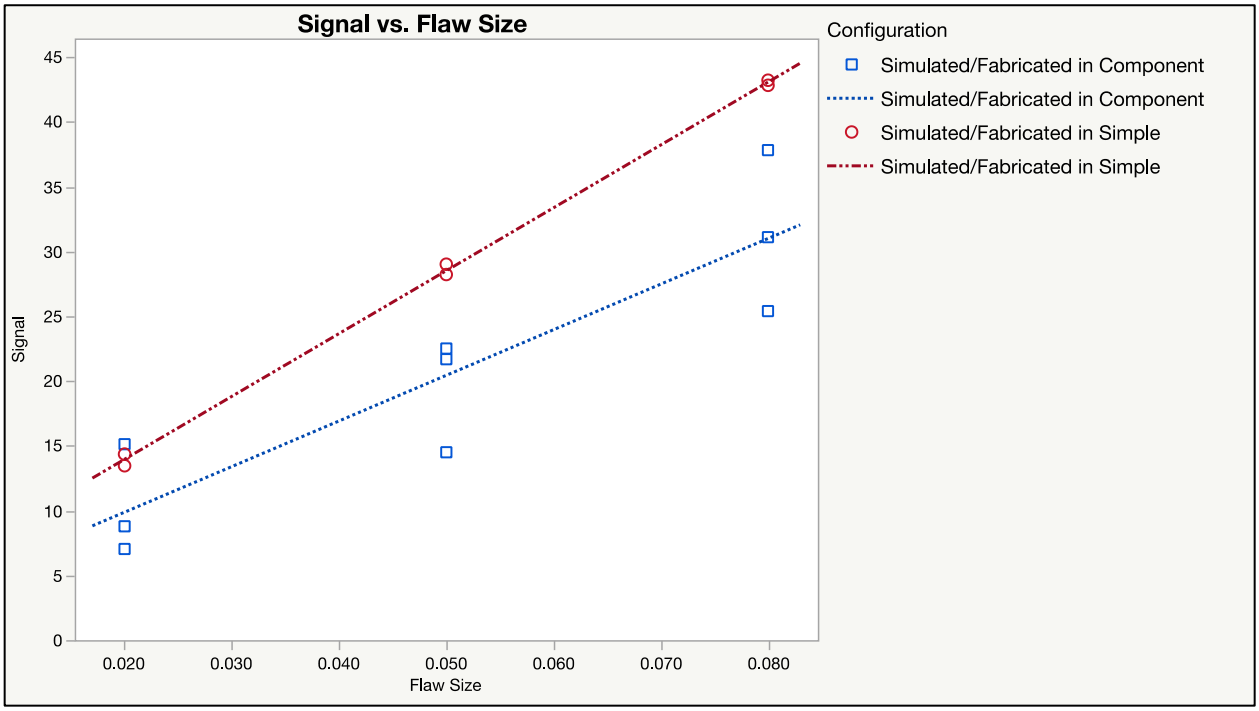


Figure 5.1-1. Plot of simulated transfer function evaluation measurements.

5.2 Modeling the Transfer Function Evaluation Measurements

A simple linear regression model in the form of $y = \beta_0 + \beta_1 x + \varepsilon$ is fit from the measurements (or their transformed variables) obtained on each specimen type, where β_0 is the intercept, β_1 is the slope, and ε is the random error that is assumed to be an independently, identically distributed, normal random variable with a mean of 0 and a variance of σ^2. From the linear regression, the model parameters are estimated as $\hat{\beta}_0$, $\hat{\beta}_1$, and s, which is the estimated standard deviation of ε for each specimen type. See Montgomery et al. (2012) for a more complete description of linear regression.

The estimated models for each specimen type are shown in Equations 1 and 2, and these expressions enable the prediction of inspection measurement signals from each specimen type as

a function of flaw size. The circumflex, or hat, notation indicates an estimated parameter or quantity.

$$\hat{y}_{NS} = \hat{\beta}_{0_{NS}} + \hat{\beta}_{1_{NS}} x_{NS} \tag{1}$$

$$\hat{y}_{NC} = \hat{\beta}_{0_{NC}} + \hat{\beta}_{1_{NC}} x_{NC} \tag{2}$$

For the transfer function evaluation measurements in Tables 5.1-1 and 5.1-2, the estimated regression model coefficients are shown in Table 5.2-1.

Table 5.2-1. Estimated Regression Coefficients from Transfer Function Evaluation Datasets

Model Parameters for Simple	Estimated Values for Simple		Model Parameters for Component	Estimated Values for Component
$\hat{\beta}_{1_{NS}}$	484		$\hat{\beta}_{1_{NC}}$	352
$\hat{\beta}_{0_{NS}}$	4.31		$\hat{\beta}_{0_{NC}}$	2.82
s_{NS}	0.45		s_{NC}	4.71

The statistical uncertainty in the model for each specimen type can be expressed with a confidence interval around the fitted linear model. The confidence interval is based on $\hat{y} \pm \Delta\hat{\mu}_{CI}$, where the formulation for $\Delta\hat{\mu}_{CI}$ is provided in Equation 3.

$$\Delta\hat{\mu}_{CI} = t_{(\alpha/2,n-2)}\, s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \tag{3}$$

where, n is the number of measurements in the dataset used to estimate the model coefficients, $t_{(\alpha/2,n-2)}$ is the percentile from a student's t-distribution for 100(1- α)% confidence interval with n-2 degrees of freedom, s is the estimated standard deviation of the signal, x_0 is a specific CIFS, and $\bar{x}$ is the average flaw size in the dataset. It is recommended to use $\alpha = 0.05$ providing a 95% confidence interval, and *t*-values are provided in Appendix A for datasets containing $n - 2 \leq 20$ inspection measurements. The term S_{xx} is the sum of squared difference between each individual flaw size and the average flaw size in the dataset, computed as:

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{4}$$

By inspection of Equation 3, it is evident that the width of the confidence interval depends on the CIFS (x_0), and it becomes wider at the extreme limits of the flaw size testing range since the squared distance from average flaw size to the CIFS increases in the $(x_0 - \bar{x})^2$ term. This confidence interval width depends on the range of the flaw sizes used in the transfer function evaluation measurement dataset, where a wider spread in flaw sizes increases S_{xx}, thereby reducing the interval width. Lastly, the interval width depends on the number of inspection measurements (n) in the dataset, where a larger number of measurements reduces the interval width. Recognizing these characteristics of Equation 3 provides information on the transfer function design parameters that are chosen by the NDE engineer and their impact on the confidence interval width.

For the transfer function evaluation measurement datasets in Tables 5.1-1 and 5.1-2, the 95% confidence intervals are shown as shaded regions around the fitted linear models in Figure 5.2-1. The confidence intervals represent the variability of the fitted models across the flaw sizes. Conceptually, a line within the shaded region of a specimen type is considered as a plausible

estimate of the signal to flaw size relationship with 95% confidence, and the dashed line in the plots represent the average model based on the specific transfer function evaluation measurement dataset. The confidence intervals represent the variability in average inspection signal level taken from multiple measurements of the same size flaw, rather than variability of individual inspection measurements that would be captured in a prediction interval.

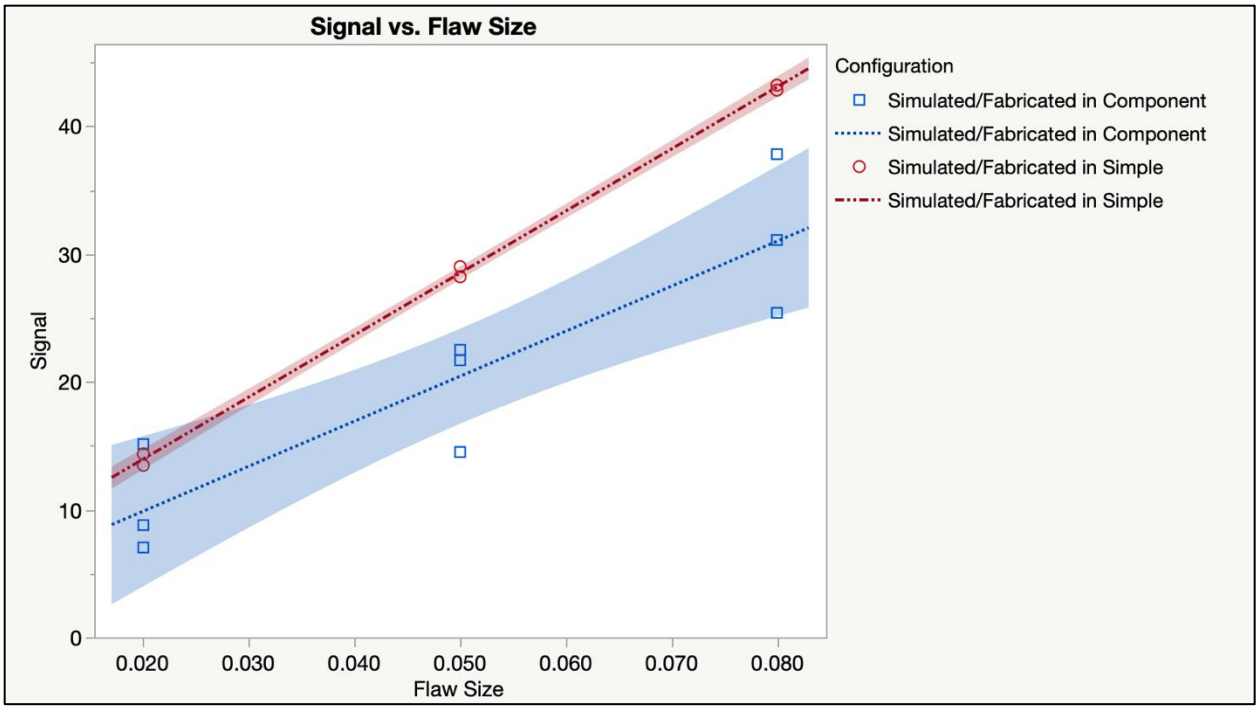


Figure 5.2-1. Plot of transfer function evaluation inspection measurements with confidence intervals around the fitted linear models.

6.0 Evaluating Similarity Between Simple and Component Specimens

After estimating the linear models and confidence intervals for both specimen types, statistical tests are performed to determine if a difference can reliably be detected between the average inspection signal level and signal variability from fabricated flaws of the same nominal size in the Component and Simple specimens. If no differences are reliably detectable, then the NDE engineer can declare similarity as discussed in NASA STD 5009B. In this case, a transfer function is not required, and this is referred to as a 1-to-1 transfer between specimen type. For this transfer, the naturally occurring CIFS desired to be detected in the flight component is equal to the simulated-induced flaw placed in a Simple specimen to perform a POD demonstration. Furthermore, if the imple specimen used in the transfer function evaluation measurement dataset is representative of the specimen configuration used to develop the Standard NDE flaw size in NASA STD 5009B, then a 1-to-1 transfer function implies that the Standard NDE flaw size can be applied to the Component specimen.

6.1 Detecting a Difference in Average Inspection Signal Between Specimen Types

The confidence intervals around the fitted lines are used to detect whether there is a difference in the average signal for the same size flaw in both specimen types. Conceptually, a hypothesis testing approach is applied by evaluating if the confidence intervals overlap at the flaw size of interest. If

the confidence intervals overlap, then no difference can be reliably detected at that flaw size, and if the confidence intervals do not overlap, then a difference in the average signal is declared. As an example, in Figure 5.2-1, at a flaw size of 0.050, a difference in the average signal is detected because the shaded confidence intervals do not overlap. However, at a flaw size of 0.020, a difference cannot be reliably detected because the intervals overlap. In regions where there is not a clear separation between the confidence intervals (e.g., just below a 0.030 flaw size) additional transfer function evaluation measurements using additional flaws may reduce the interval width. This simplified hypothesis testing approach is considered sufficient since the transfer function evaluation measurements are not being used to assess POD directly, instead they are used to inform the need for a subsequent POD demonstration.

Numerically, the upper and lower bounds of the confidence intervals are computed from Equation 3 at a flaw size of interest. For the example data in Tables 5.1-1 and 5.1-2, the upper and lower bounds of the confidence intervals for a 0.065 flaw size are shown in Table 6.1-1.

Table 6.1-1. Estimated Model Parameters and Confidence Intervals for Simple and Component Specimens at a 0.065 Flaw Size

Estimated model parameter	Simple Specimen		Estimated model parameter	Component Specimen
$\hat{y}_{NS}$	35.8		$\hat{y}_{NC}$	25.7
t-value$_{NS}$	2.78		t-value$_{NC}$	2.36
$\Delta\hat{\mu}_{CI_{NS}}$	0.60		$\Delta\hat{\mu}_{CI_{NC}}$	4.91
$\hat{\mu}_{U_{NS}}$	36.4		$\hat{\mu}_{U_{NC}}$	30.6
$\hat{\mu}_{L_{NS}}$	35.2		$\hat{\mu}_{L_{NC}}$	20.8

By inspection, the lower confidence bound ($\hat{\mu}_{L_{NS}}$) for the simulated-fabricated flaw in the Simple specimen exceeds the upper confidence bound ($\hat{\mu}_{U_{NC}}$) for the simulated-fabricated flaw in the Component specimen. Thus, the confidence intervals do not overlap, and a difference in the average signal is detected.

In the example data, the fitted model of the signal is lower from the Component specimen for all flaw sizes within the range of the transfer function evaluation measurements. However, the approach presented remains valid if the signal is higher (i.e., amplified). If the Component specimen signal is greater than the Simple specimen signal, the comparison would be between the upper confidence bound for the Simple specimen and the lower confidence bound for the Component specimen.

6.2 Detecting a Difference in the Signal Variability Between Specimen Types

The estimated signal variability from each specimen type is used to determine if a difference in the signal variability between specimen types can be reliably detected. Since fabricated flaws are utilized in the transfer function evaluation measurement datasets, it is assumed that differences in signal variability are primarily due to specimen features (e.g., effects of surface finish or inspection access restrictions). The evaluation of similarity is based on signal standard deviation estimates as pure-error estimates from the replicated flaws, or using the residuals from the regression modeling. There will be negligible difference in these two estimating approaches if the regression model fits

well. Following best practice in regression modeling, evaluating the goodness of fit is recommended using multiple quantitative metrics and graphical diagnostics to confirm that the regression model is appropriate, see Montgomery et al., (2012). While pure-error estimates are preferred, the regression residual root mean squared error (RMSE) is typically readily available from common analysis software packages. With the NDE engineer in mind, the residual RMSE from the regression is considered sufficient for the purpose of this test. The RMSE represents the average variability across the range of flaw sizes in the transfer function evaluation measurement datasets, and therefore it is used in the following analyses as being independent of flaw size.

A hypothesis testing approach is employed to determine if a difference in the signal variability can be reliably detected, where the null (H_0) and alternative (H_1) hypotheses are based on ratios of the variances from each specimen type as:

$$H_0: \frac{\sigma_{NC}^2}{\sigma_{NS}^2} = 1 \tag{5}$$

$$H_1: \frac{\sigma_{NC}^2}{\sigma_{NS}^2} > 1 \tag{6}$$

where σ^2 represents the true signal variability. To conduct the test, an observed test statistic (f_{obs}) is computed from the regression RMSE standard deviations from the transfer function evaluation measurements as:

$$f_{obs} = \left(\frac{s_{NC}}{s_{NS}}\right)^2 \tag{7}$$

As the ratio of the estimated signal variabilities approach 1, then the NDE engineer would not be able to reliably detect a difference in the variability. As the ratio increases, the evidence becomes stronger that there is a detectable difference. The test statistic is compared to a value from an *F*-distribution, $F_{(\alpha, n_{NC}-2, n_{NS}-2)}$, which is a percentile value associated with 100(1- α)% probability with $(n_{NC} - 2)$ and $(n_{NS} - 2)$ degrees of freedom. This *F*-value is known as a cut-off value for the hypothesis test. In this expression, n_{NC} and n_{NS} are the number of inspection measurements from the Component and Simple specimens, respectively. It is recommended to use $\alpha = 0.05$, and *F*-values are provided in Appendix B for datasets containing $(n - 2) \leq 20$ inspection measurements. If the test statistic exceeds the cut-off value from the *F*-distribution, such that

$$f_{obs} > F_{(\alpha, n_{NC}-2, n_{NS}-2)} \tag{8}$$

then a difference in variability between the specimen types is declared. This formulation assumes that the variability in the Component specimen will typically be larger than the variability in the Simple specimen. In cases where the variability in the Simple specimen exceeds that of the Component specimen, it is suggested that the inspection data and procedures be reviewed to confirm this result.

The *F*-value table in Appendix B shows that it is more difficult to detect a small difference in variability with few inspection measurements in the transfer function evaluation measurement datasets. At the minimum recommended number of measurements of $n_{NC} = 9$ and $n_{NS} = 6$, the squared ratio of the signal variability must exceed 6.1 for a difference in the signal variability to be reliably detected (i.e., the signal standard deviation measured in the Component specimen needs to be greater than $\sqrt{6.1} = 2.5$ times the signal standard deviation in the Simple specimen).

For the transfer function evaluation measurement data in Tables 5.1-1 and 5.1-2, the test statistic and *F*-distribution cut-off value are computed in Table 6.2-1.

Table 6.2-1. Signal Variability Test Statistic and F-distribution Cut-off Value

s_{NC}^2/s_{NS}^2	107.6
Cut-off value to Exceed	6.1

In this example, the test statistic exceeds the cut-off value, and therefore, a difference in the variability is detected and the development of a transfer function based on variability is needed.

7.0 Estimating Transfer Functions when Similarity is not declared

If similarity is not found, then transfer functions are needed. In the following, transfer function for the average signal level and an adjustment based on differences in variability are presented for the Forward and Inverse case. For each application case, suggested requirements are provided to perform a Special NDE demonstration

7.1 Forward Case

If a difference in the average signal level and/or the variability is detected between specimen types, then similarity cannot be claimed based on the transfer function evaluation measurements. Thus, a transfer function approach is needed. Recall that in the Forward case, a naturally occurring CIFS (x_t) in a flight component is provided and a representative simulated-induced flaw size is sought to be specified in a Simple specimen for Special NDE POD demonstration testing by all inspectors to be qualified. As an example of the Forward case, consider that the detection of a naturally occurring CIFS of $x_t = 0.065$ in the flight component is specified by fracture analysis. What is the simulated-induced demonstration flaw size (d) in a Simple specimen for POD demonstration? To estimate d, adjustments to the average signal level and/or signal variability are performed via transfer functions resulting in $\hat{d}$.

7.1.1 Forward Case: Transferring the Average Signal Level

To transfer the average signal level from Component to Simple specimens, x_{NC} is replaced with x_t in Equation 2 and the predicted signal level ($\hat{y}_t$) from a flaw of size x_t in a Component specimen is predicted as:

$$\hat{y}_t = \hat{\beta}_{0_{NC}} + \hat{\beta}_{1_{NC}} x_t \tag{9}$$

Then, x_{NS} is replaced by $\hat{d}_{mean}$, which is the demonstration flaw size sought, and $\hat{y}_{NS}$ is replaced by $\hat{y}_t$ in Equation 1.

$$\hat{y}_t = \hat{\beta}_{0_{NS}} + \hat{\beta}_{1_{NS}} \hat{d}_{mean} \tag{10}$$

Setting these two equations equal to each other and rearranging to estimate the demonstration flaw size $\hat{d}_{mean}$, provides the expression:

$$\hat{d}_{mean} = \frac{\left(\hat{\beta}_{0_{NC}} + \hat{\beta}_{1_{NC}} x_t\right) - \hat{\beta}_{0_{NS}}}{\hat{\beta}_{1_{NS}}} \tag{11}$$

This equation estimates the demonstration flaw size using the coefficients of the linear models from both specimen types and the given CIFS to be detected in the flight component.

Using the example datasets from Tables 5.1-1 and 5.1-2 and $x_t = 0.065$, Figure 7.1-1 graphically illustrates the process where the POD demonstration flaw size is computed as:

$$\hat{d}_{mean} = \frac{(2.82+352(0.065))-4.31}{484} = 0.044 \tag{12}$$

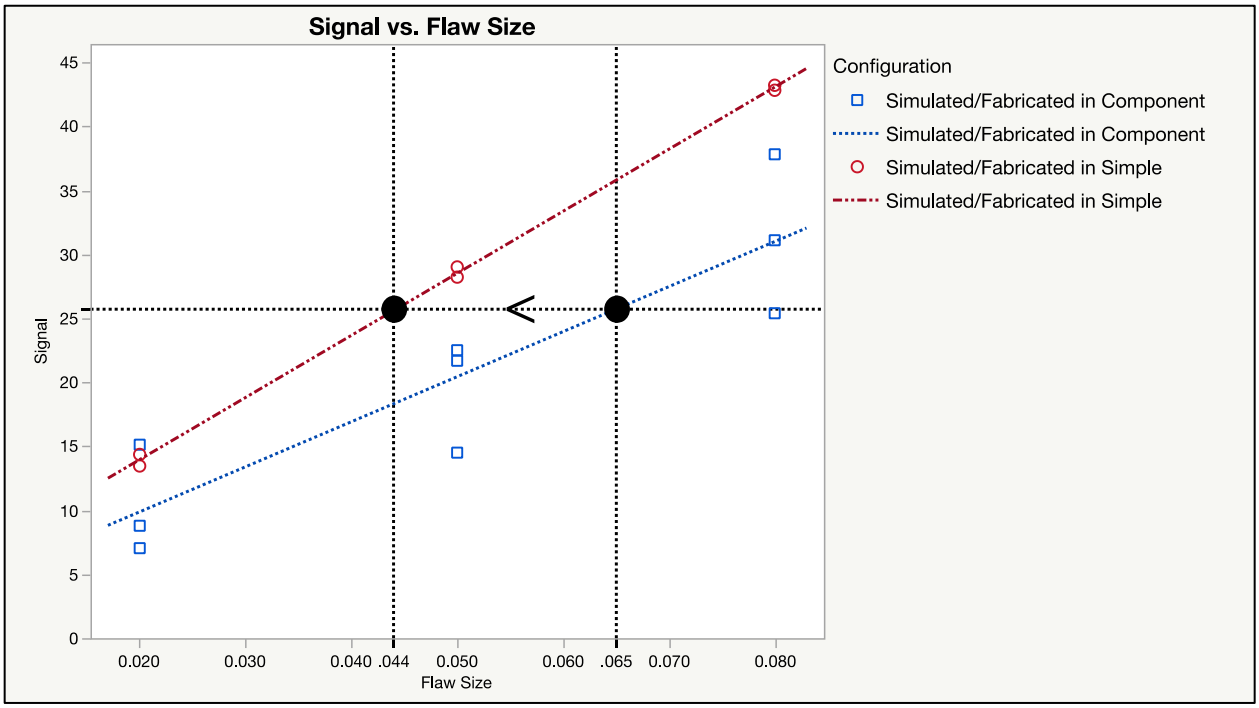


Figure 7.1-1. Plot illustrating transfer of average signal in forward case.

7.1.2 Forward Case: Transferring the Signal Variability

If a difference in the signal variability is detected between the Component and Simple specimen, then the demonstration flaw size based on the difference in the average signal level ($\hat{d}_{mean}$) is adjusted by $\Delta\hat{d}_{var}$. If no difference in the average signal level was detected, where $\hat{d}_{mean} = x_t$, then an adjustment due to variability can be considered. The procedure converts the standard deviation of the signal variability (s) to an adjustment in flaw size ($\Delta\hat{x}_{var}$) using the estimated slopes from the linear models for each specimen type.

$$\Delta\hat{x}_{var_{NS}} = \frac{s_{NS}}{\hat{\beta}_{1_{NS}}} \tag{13}$$

$$\Delta\hat{x}_{var_{NC}} = \frac{s_{NC}}{\hat{\beta}_{1_{NC}}} \tag{14}$$

Recognizing that the baseline signal variability from the Simple specimen is included in the signal variability in the Component, the adjustment is based on the increase in variability in the Component specimen as:

$$\Delta\hat{d}_{var} = \Delta\hat{x}_{var_{NC}} - \Delta\hat{x}_{var_{NS}} \tag{15}$$

Using the example datasets from Tables 5.1-1 and 5.1-2, the adjustment to the demonstration flaw size is computed.

$$\Delta\hat{d}_{var} = \frac{s_{NC}}{\hat{\beta}_{1_{NC}}} - \frac{s_{NS}}{\hat{\beta}_{1_{NS}}} = \frac{4.71}{352} - \frac{0.45}{484} = 0.013 - 0.001 = 0.012 \tag{16}$$

7.1.3 Forward Case: Reporting the Final Results

Combining the transfer of the average signal and adjusting for increased signal variability, the demonstration flaw size is computed.

$$\hat{d} = \hat{d}_{mean} - \Delta\hat{d}_{var} \tag{17}$$

If no difference in the average signal level was detected then $\hat{d}_{mean}$, is replaced by x_t. Using the example datasets from Tables 5.1-1 and 5.1-2, the demonstration flaw size can be computed.

$$\hat{d} = 0.044 - 0.012 = 0.032 \tag{18}$$

In this example, a Special NDE POD demonstration would be performed using simulated-induced flaws of 0.032 or smaller, in a Simple specimen that is the same configuration as the Simple specimen used in the transfer function evaluation measurement dataset by each inspector to be qualified.

7.2 Inverse Case

The Inverse case follows the same approach discussed for the Forward case to first test if a difference in the average signal level and/or the variability are detected between specimen types. If similarity cannot be claimed based on the transfer function evaluation measurements, then an Inverse transfer function could be developed. Recall for the Inverse case, a 90/95 flaw size from a POD study exists using simulated-induced flaws in a Simple specimen, and the goal is to estimate a representative naturally occurring flaw size in a flight component. As an Inverse example, consider a simulated-induced flaw size (d) of 0.040 that has a demonstrated 90/95 POD in a Simple specimen. What is the estimated representative naturally occurring flaw size ($\hat{c}$) in the flight component?

7.2.1 Inverse Case: Transferring the Average Signal Level

The Simple specimen in the transfer function evaluation measurement dataset is assumed to be the same as the specimen used in the Standard or Special NDE POD study that produced the POD flaw size to be transferred. If a Special NDE flaw size is being transferred, then the representative flaw size in the flight component is applicable to inspectors who were qualified through the original Special NDE demonstration. If a Standard NDE flaw size is transferred, then the representative flaw size in the flight component is considered applicable to Level II qualified inspectors as specified by NASA STD 5009B, as would be the case for the application of Standard NDE flaw sizes.

For the example scenario, testing for detectable differences in the average signal and variability are performed for a flaw size of 0.040 following the same process previously described. While detailed calculations of the confidence intervals at 0.040 flaw size are not provided for brevity, by visual inspection of Figure 5.2-1, it is evident that the confidence intervals do not overlap at a flaw size of 0.040, and therefore a difference in the average signal level can be detected. The test for the difference in the signal variability, which does not depend on flaw size, is identical to the previous analysis, and a difference is detected. Therefore, an inverse transfer of the average signal and signal variability could be developed.

If a difference in the average signal at the flaw size of interest is detected between the Component and Simple specimen, then x_{NS} is replaced with $\hat{d}$ in Equation 1 and the signal level ($\hat{y}_d$) in a Simple specimen at that flaw size is predicted by:

$$\hat{y}_d = \hat{\beta}_{0_{NS}} + \hat{\beta}_{1_{NS}}\hat{d} \tag{19}$$

Then, x_{NC} is replaced by $\hat{c}_{mean}$ and $\hat{y}_{NC}$ is replaced by $\hat{y}_d$ in Equation 2.

$$\hat{y}_d = \hat{\beta}_{0_{NC}} + \hat{\beta}_{1_{NC}}\hat{c}_{mean} \tag{20}$$

Setting these two equations equal to each other and rearranging to estimate the flight component flaw size $\hat{c}_{mean}$, provides the expression:

$$\hat{c}_{mean} = \frac{\left(\hat{\beta}_{0_{NS}} + \hat{\beta}_{1_{NS}}\hat{d}\right) - \hat{\beta}_{0_{NC}}}{\hat{\beta}_{1_{NC}}} \tag{21}$$

This equation estimates the average flaw size ($\hat{c}_{mean}$) in the flight component as a function of the linear model coefficients and the previously demonstrated POD flaw size from a Simple specimen.

Using the example datasets from Tables 5.1-1 and 5.1-2 and $\hat{d} = 0.040$, the estimated naturally occurring flaw size can be computed, and Figure 7.2-1 graphically illustrates the process.

$$\hat{c}_{mean} = \frac{(4.31 + 484(0.040)) - 2.82}{352} = 0.059 \tag{22}$$

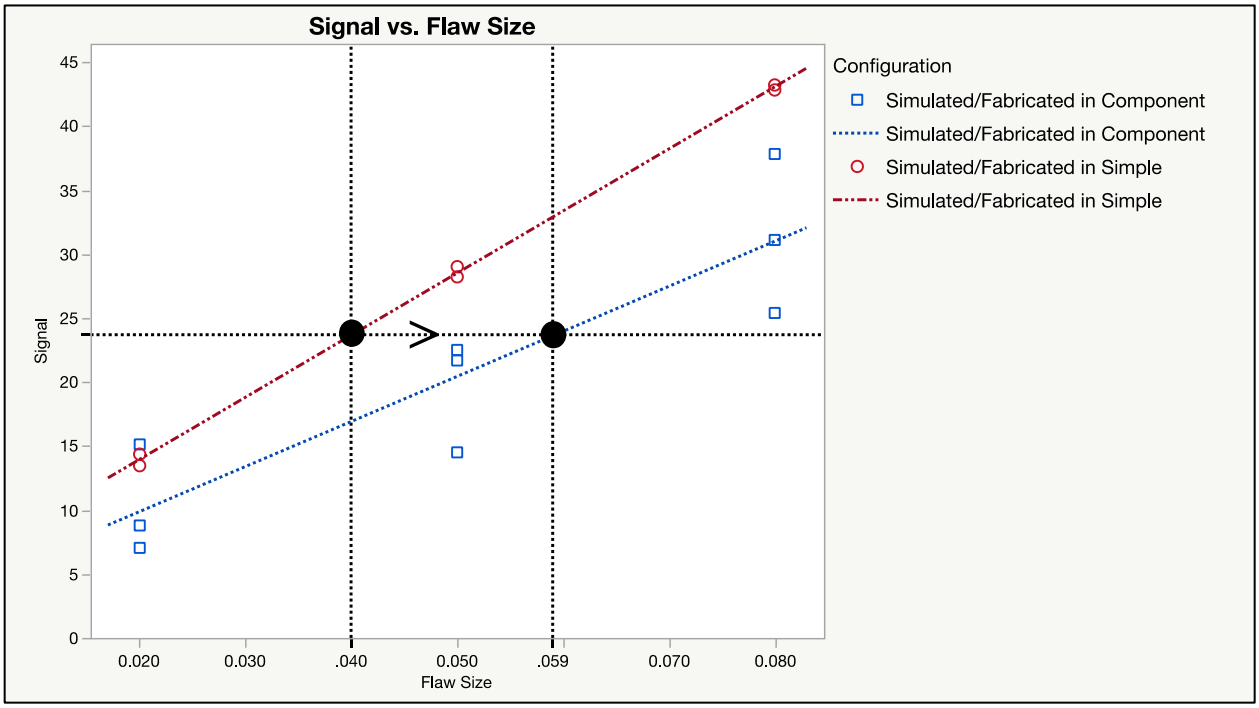


Figure 7.2-1. Plot illustrating transfer of average signal in inverse case.

7.2.2 Inverse Case: Transferring the Signal Variability

If a difference in the signal variability is detected between the Component and Simple specimen from the transfer function evaluation measurement data, then the estimated naturally occurring flaw size based on the difference in the average signal level ($\hat{c}_{mean}$) is adjusted by $\Delta\hat{c}_{var}$. If no difference in the average signal level was detected, where $\hat{c}_{mean} = \hat{d}$, then an adjustment due to

variability can be considered. Similar to the Forward case, the procedure converts the standard deviation of the signal variability (s) to an adjustment in flaw size ($\Delta\hat{x}_{var}$) using the estimated slopes from the linear models for each specimen type.

$$\Delta\hat{x}_{var_{NS}} = \frac{s_{NS}}{\hat{\beta}_{1_{NS}}} \tag{23}$$

$$\Delta\hat{x}_{var_{NC}} = \frac{s_{NC}}{\hat{\beta}_{1_{NC}}} \tag{24}$$

While Equations 23 and 24 are identical to 13 and 14, they are repeated and given unique numbers for the reader's convenience. Assuming that the signal variability is larger in the Component specimen, the estimated Simple adjustment is subtracted from the Component adjustment in Equation 25 to increase in the average flaw size detectable in the flight component.

$$\Delta\hat{c}_{var} = \Delta\hat{x}_{var_{NC}} - \Delta\hat{x}_{var_{NS}} \tag{25}$$

Using the example datasets from Tables 5.1-1 and 5.1-2, the adjustment to the estimated naturally occurring flaw size is computed.

$$\Delta\hat{c}_{var} = \frac{s_{NC}}{\hat{\beta}_{1_{NC}}} - \frac{s_{NS}}{\hat{\beta}_{1_{NS}}} = \frac{4.71}{352} - \frac{0.45}{484} = 0.013 - 0.001 = 0.012 \tag{26}$$

7.2.3 Inverse Case: Reporting the Results

Combining the transfer of the average signal and adjusting for signal variability, the estimated flaw size in the flight component is computed as:

$$\hat{c} = \hat{c}_{mean} + \Delta\hat{c}_{var} \tag{27}$$

Using the example datasets from Tables 5.1-1 and 5.1-2, the flight component flaw size can be computed by:

$$\hat{c} = 0.059 + 0.012 = 0.071 \tag{28}$$

In this example, a 0.040 flaw size from an existing POD study with a Simple specimen is estimated to be representative of a 0.071 flaw size in the flight component.

8.0 Protocols for Special NDE Demonstration Based on Transfer Functions

In this section, suggested protocols for performing a Special NDE demonstration, as required by NASA STD 5009B, are presented for the three possible transfer function outcomes; (1) Similarity is accepted and a 1-to-1 transfer function is assumed, (2) Similarity is not accepted and an induced flaw size is specified for demonstration in the Forward case, or (3) Similarity is not accepted and the naturally occurring flaw size is estimated from a previous POD demonstration or from the Standard NDE table in NASA STD 5009B. For all three outcomes, individual inspector certification is based on a combination of a POD demonstration and a technique consistency demonstration. The POD demonstration uses simulated-induced flaws that are representative of the flaw-to-flaw variability and flaw morphology that is anticipated in the flight component. The technique consistency demonstration uses the Component specimens that simulate the flight hardware inspection geometry, material, and inspection access and were used to develop the transfer function. This combination of an inspector demonstration of POD with induced flaws and demonstration of their ability to detect fabricated flaws in the Component specimen provides the

most comprehensive assessment of inspection capability when there is an insufficient number of independently verified naturally occurring flaws in flight components.

8.1 Transfer Function Outcome 1 – Similarity Accepted, and 1-to-1 Transfer Function is Assumed

If similarity to a previous Special NDE or Standard NDE flaw size is accepted, then it is assumed that each inspector will have performed the Special NDE POD demonstration of that flaw size or is qualified to perform Standard NDE inspections.

In addition, it is recommended that each inspector demonstrates their consistency with the signal response model of the Component specimen containing fabricated flaws. Ideally, the same Component specimen(s) used to evaluate similarity would be used in this demonstration. If that is not possible, then specimens used are to be documented to be equivalent by the NDE engineer. The inspections are to be performed unblinded (i.e., where the inspector has knowledge of the flaw location) for each fabricated flaw and the signal response is recorded. Using the approach described in Section 5.2 (Modeling the Transfer Function Evaluation measurements), a signal versus flaw size model is estimated and compared to the confidence intervals estimated during the evaluation of similarity in Section 6.1. If the individual inspector model is within the confidence intervals, then it is considered consistent, and the inspector has passed this certification step. If the individual inspector model is not consistent, then a diagnostic phase is initiated to determine the source of the discrepancy which may include calibration, standardization, and/or technique. This diagnostic phase continues until the inspection process achieves consistency.

8.2 Transfer Function Outcome 2 – Similarity Not Accepted, and Forward Transfer Function is Applied

If a Forward transfer function is applied, the first certification step requires each individual inspector to perform a POD demonstration with induced flaws in a Simple specimen. Acceptable methods for this POD demonstration are specified in NASA STD 5009B and include the point estimate method and MIL-HDBK-1823A (2009). In addition, the limited sample POD method described by Koshti et al. (2021b) could be considered acceptable by a FCB on a case-by-case basis.

In addition, it is recommended that each inspector demonstrates their capability with the signal response model of the Component specimen containing fabricated flaws. Ideally, the same Component specimen(s) used to evaluate similarity would be used in this demonstration. If that is not possible, specimens documented to be equivalent by the NDE engineer are required. The inspections are to be performed blind, meaning without the inspector's prior knowledge of the flaw locations. The inspector is required to find every fabricated flaw and report the signal response. Using the approach described in Section 5.2, a signal versus flaw size model is estimated and compared to the confidence intervals estimated during the evaluation of similarity in Section 6.1. If the individual inspector model is within the confidence intervals, then it is considered consistent, and the inspector has passed this certification step. If the individual inspector model is not consistent, then a diagnostic phase is initiated to determine the source of the discrepancy, which may include calibration, standardization, and/or technique.

8.3 Transfer Function Outcome 3 – Similarity Not Accepted, and Inverse Transfer Function is Applied

If an Inverse transfer function is applied, it is assumed that each inspector will have performed the Special NDE POD demonstration of that flaw size or is qualified to perform Standard NDE inspections.

In addition, it is recommended that each inspector demonstrates their consistency with the signal response model of the Component specimen containing fabricated flaws. Ideally, the same Component specimen(s) used to evaluate similarity would be used in this demonstration. If that is not possible, specimens documented to be equivalent by the NDE engineer are required. The inspections are to be performed blind. The inspector is required to find every fabricated flaw and report the signal response. Using the approach described in Section 5.2, a signal versus flaw size model is estimated and compared to the confidence intervals. If the individual inspector model is within the confidence intervals estimated during the evaluation of similarity in Section 6.1, then it is considered consistent, and the inspector has passed this certification step. If the individual inspector model is not consistent, then a diagnostic phase is initiated to determine the source of the discrepancy, which may include calibration, standardization, and/or technique. This diagnostic phase continues until the inspection process achieves consistency.

9.0 Overview of Transfer Function Steps

This section provides a condensed summary of the transfer function process described in this guidebook to assist NDE engineers in the design, execution, and analysis of a transfer function study. Formulae and quantitative criteria for each step are not repeated for brevity, and they can be found in the relevant sections in this guidebook.

1. Determine if a Forward or Inverse transfer is required based on the application.
    a. Forward transfer begins with the naturally occurring flaw size of interest to be detected in the flight component, and it specifies a simulated-induced flaw in a Simple specimen for Special NDE POD demonstration (or acceptance of a Standard flaw size if the targeted simulated-induced flaw is greater than or equal to the Standard flaw size for one of the Standard NDE methods specified in NASA STD 5009B).
    b. Inverse transfer begins with an existing 90/95 flaw size from the Standard NDE table of NASA STD 5009B or a previous Special NDE demonstration, and it predicts the equivalent flaw size in the flight component.
2. Define a range of simulated-fabricated flaw sizes that cover the CIFS of interest in the flight component.
3. Define the Simple specimen geometry (e.g., flat plate, round bar) and the Component specimen (e.g., representative flight geometry, material processing, and inspection access restrictions).
4. Fabricate Simple and Component specimens with fabricated flaws (e.g., EDM notches)
5. Perform inspection measurements using the same NDE method, setup (i.e., instrument and standardization/calibration) and procedure with the same inspector for both specimens.
6. Define the appropriate flaw size parameter to be modeled (e.g., length, depth, area, aspect ratio).

7. Model the signal versus flaw size relationships of the Simple and Component specimens.
8. Evaluate Similarity by determining whether a difference in the mean signal level and/or variability is reliably detectable within the inspection measurement variability.
    a. If no difference is detectable, then similarity can be claimed (i.e., a 1-to-1 transfer function), and the CIFS for the flight component would be specified in a Simple specimen for Special NDE POD demonstration.
    b. If a difference in the mean and/or variaibility is detected between specimen types, then transfer function(s) are developed.
9. Calculate the flaw size that produces the same average signal in both specimens. For a Forward transfer, this is a transfer from the Component to the Simple specimen. For an Inverse transfer, this is a transfer from the Simple to the Component specimen.
10. Calculate an adjustment based on the difference in variability and adjust the average flaw size. For a Forward transfer, the average flaw size is decreased by the variaiblity adjustment, and for an Inverse transfer, the average flaw size is increased by the variability adjustment.
11. Report the transferred simulated-induced (e.g., fatigue crack) flaw size that is recommended for POD demonstration for the Forward case, or the predicted flaw size in the flight component for the Inverse case.
12. Perform Special NDE POD demonstration as per NASA STD 5009B (unless the target simulated induced flaw is greater than a NASA STD flaw size for a Standard NDE Method identified in NASA STD 5009B), and perform the additional transfer function specific demonstration with the component specimen. The additional Special NDE criteria and protocol differ depending on the finding of Similarity or if a Forward or Inverse transfer is developed.

10.0 Conclusion

NASA requires that NDE methods and inspectors used for fracture critical hardware inspections on human spacecraft provide 90/95 POD. However, NDE POD demonstrations are time consuming, and in many cases, it is impractical to perform the POD demonstrations on the flight component of interest for inspection. Thus, there is a strong desire to generalize the results from a given POD demonstration test to other similar inspection scenarios, and provide a practical approach to demonstrating POD when the demonstration testing cannot be performed on the specific component of interest. This guidebook provides a general approach to evaluate similarity and develop transfer functions that is considered intuitive, cost-effective, and implementable by NDE engineers. This methodology overcomes gaps in the NDE literature on transfer functions, and it is specifically tailored to comply with NASA STD 5009B for fracture-critical metallic components in human-rated aerospace flight systems. The approach is based on using simulated-fabricated flaws (e.g., EDM notches) in Simple and Component specimens, which makes specimen preparation more feasible. The rationale for utilizing a specimen-type transfer compared to a flaw-type transfer was presented and it guards against a nonconservative POD evaluation. The quantitative evaluation of similarity is a significant contribution to support the utilization of provisions in NASA STD 5009B for transferability and similarity. Specific criteria for the design, execution, and analysis of a transfer function study are provided. This method can be readily tailored for specific NDE methods and applications by a NDE engineer. Guidance is provided for demonstrating POD capability with simulated-induced flaws (e.g., fatigue cracks) in Simple specimens and additional Component demonstrations ensure consistency with the transfer function

development. A simple, simulated example is used throughout the presentation of the method, and case studies are provided in Appendix C to illustrate the method in practice. As with all NDE inspections, the statistical experiment design and analysis approach presented is to complement, not replace, the experience and knowledge of the NDE engineer. Acceptance and implementation of this methodology has the potential to significantly reduce risk and schedule impact on NASA programs for qualifying NDE methods and inspectors. Additionally, codifying the process to be used for transfer functions such that a consistent methodology is used should result in increased safety and reliability for NASA programs.

11.0 References

Bishop, C. R. (1973): "Nondestructive Evaluation of Fatigue Cracks," NASA Contractor Report NAS9-14000.

Bode, M. D.; Smith, B. B.; and Spencer, F. W. (2008): "Model-Assisted Probability of Detection Validation of Eddy Current Inspection for Cracks in Lap Splice Fastener Holes," *Aging Aircraft 2008 Conference*, April 21–24, 2008, Phoenix, AZ, SAND2008=2603C.

Bode, M. D.; Newcomer, J.; and Fitchett, S. (2012): "Transfer function model-assisted probability of detection for lap joint multi site damage detection," *AIP Conference Proceedings 1430*, 1749 (2012), Review of Progress in Quantitative Nondestructive Evaluation, Volume 31.

Harding, C.; Hugo, G.; and Bowles, S. (2007): "Model-Assisted Probability of Detection Validation of Automated Ultrasonic Scanning for Crack Detection at Fastener Holes," *Proceedings of the 10th Joint FAA/DoD/NASA Conference on Aging Aircraft*, Palm Springs, CA, April 16-19, 2007.

Harding, C.; Hugo, G.; and Bowles, S. (2008): "Application of Model-Assisted POD Using a Transfer Function Approach," *35th Annual Review of Quantitative Nondestructive Evaluation*, Volume 28, pp. 1792–1799. July 22–25, 2008, Chicago, Illinois. American Institute of Physics, Melville, New York. AIP Conf. Proc. 1096.

Harding, C. A.; Hugo, G. R.; and Bowles, S. J. (2009): "Application of Model-Assisted POD Using a Transfer Function Approach," *AIP Conference Proceedings 1096*, 1792, 2009.

Jones, K. J.; Brausch, J.; Fong, W.; and Harris, B. (2015): "Probing the Future: Better F-16 Inspections Using Conformal Eddy Current Inspection Tools," *28th International Committee on Aeronautical Fatigue and Structural Integrity*, June 3–5.

Koshti, A. (2021a): "Assessment of flaw detectability using transfer function," Proc. SPIE 11592, Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XV, 115920O, March 22, 2021.

Koshti, A.; Parker, P.; Forsyth, D.; Suits, M.; Walker, J.; and Prosser, W. (2021b): "Guidebook for Limited Sample Probability of Detection (LS-POD) Demonstration for Single-Hit Nondestructive (NDE) Methods," NASA TM-20210018515.

Lindgren, E.; Forsyth, D.; Aldrin, J.; and Spencer, F.: "Probability of Detection," ASM Handbook Volume 17, Nondestructive Evaluation of Materials, 2018.

Meyer, R. M.; Crawford, S. L.; Lareau, J. P.; and Anderson, M. T. (2014): "Review of Literature for Model Assisted Probability of Detection," Pacific Northwest National Laboratory, PNNL-23714, September 2014.

MIL-HDBK-1823A (2009): "Nondestructive Evaluation System Reliability Assessment."

Montgomery; Peck; and Vining (2012): Introduction to Linear Regression Analysis, 5th Edition, John Wiley & Sons, Hoboken.

NASA Standard 5009B (2019): "Nondestructive Evaluation Requirements for Fracture-Critical Metallic Components."

Thompson, R. B.; Brasche, L. J.; Forsyth, D.; Lindgren, E.; Swindell, P.; and Winfree, W.: "Recent Advances in Model-Assisted Probability of Detection," 4th European-American Workshop on Reliability of NDE, June24-26, 2009.

Appendix A. t-distribution Values for 95% Confidence Interval Around the Fitted Line

$n-2$	t-value
4	2.78
5	2.57
6	2.45
7	2.36
8	2.31
9	2.26
10	2.23
11	2.20
12	2.18
13	2.16
14	2.14
15	2.13
16	2.12
17	2.11
18	2.10
19	2.09
20	2.09

Appendix B. F-distribution Values used in Testing the Difference in Signal Variability, α = 0.05

	$n_{NC} - 2$													
$n_{NS} - 2$	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
4	6.1	6.0	6.0	6.0	5.9	5.9	5.9	5.9	5.9	5.8	5.8	5.8	5.8	5.8
5	4.9	4.8	4.8	4.7	4.7	4.7	4.7	4.6	4.6	4.6	4.6	4.6	4.6	4.6
6	4.2	4.1	4.1	4.1	4.0	4.0	4.0	4.0	3.9	3.9	3.9	3.9	3.9	3.9
7	3.8	3.7	3.7	3.6	3.6	3.6	3.6	3.5	3.5	3.5	3.5	3.5	3.5	3.4
8	3.5	3.4	3.4	3.3	3.3	3.3	3.3	3.2	3.2	3.2	3.2	3.2	3.2	3.2
9	3.3	3.2	3.2	3.1	3.1	3.1	3.0	3.0	3.0	3.0	3.0	3.0	2.9	2.9
10	3.1	3.1	3.0	3.0	2.9	2.9	2.9	2.9	2.8	2.8	2.8	2.8	2.8	2.8
11	3.0	2.9	2.9	2.9	2.8	2.8	2.8	2.7	2.7	2.7	2.7	2.7	2.7	2.6
12	2.9	2.8	2.8	2.8	2.7	2.7	2.7	2.6	2.6	2.6	2.6	2.6	2.6	2.5
13	2.8	2.8	2.7	2.7	2.6	2.6	2.6	2.6	2.5	2.5	2.5	2.5	2.5	2.5
14	2.8	2.7	2.6	2.6	2.6	2.5	2.5	2.5	2.5	2.4	2.4	2.4	2.4	2.4
15	2.7	2.6	2.6	2.5	2.5	2.5	2.4	2.4	2.4	2.4	2.4	2.4	2.3	2.3
16	2.7	2.6	2.5	2.5	2.5	2.4	2.4	2.4	2.4	2.3	2.3	2.3	2.3	2.3
17	2.6	2.5	2.5	2.4	2.4	2.4	2.4	2.3	2.3	2.3	2.3	2.3	2.2	2.2
18	2.6	2.5	2.5	2.4	2.4	2.3	2.3	2.3	2.3	2.2	2.2	2.2	2.2	2.2
19	2.5	2.5	2.4	2.4	2.3	2.3	2.3	2.3	2.2	2.2	2.2	2.2	2.2	2.2
20	2.5	2.4	2.4	2.3	2.3	2.3	2.2	2.2	2.2	2.2	2.2	2.2	2.1	2.1

Appendix C. Case Studies in Evaluation Similarity and Developing Transfer Functions

Example 1 – Similarity in Average Signal and Variability

The following example is based on a presentation by Smith et al. (2009) to the model-assisted probability of detection (MAPOD) Working Group and invovles the inspection of complex aircraft geometery compared to simple flat plates, where both specimens have electrical discharge machining (EDM) notches as fabricated flaws, shown in Figures C.1 and C.2. The original data were not available, therefore a simulation was generated for illustration, courtesy of Meeker (2021). The flaw size units are normalized and are non-dimensional.

For this example, the criticial initial flaw size of interest is 0.40 in the Component specimen. Ten fabricated flaws ranging in size were inspected from each specimen type, and the transfer function evaluation dataset is provided in Table C.1. A plot of signal versus flaw size is shown in Figure C.3 that distinguishes each specimen type by marker, and the CIFS of 0.4 is indicated with a vertical dashed line. The estimated linear model coefficients and RMSE are provided for each specimen type on the plot. The shaded regions are 95% confidence intervals on the regression line.

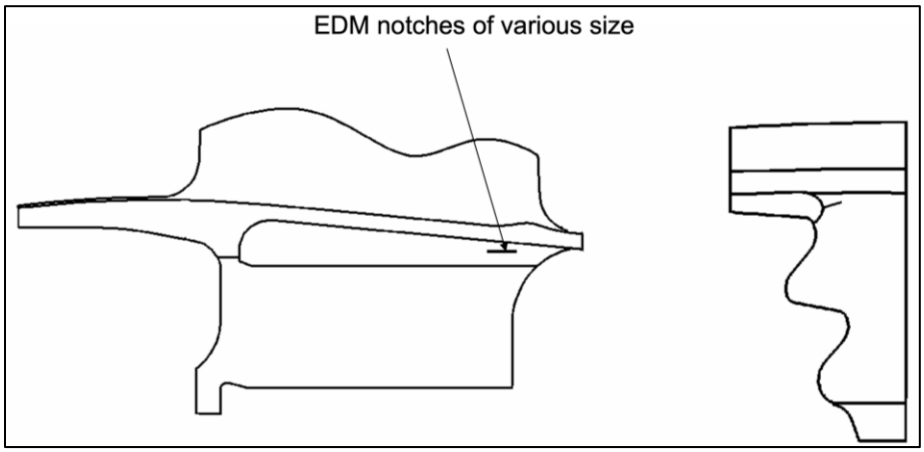


Figure C.1. Example 1 - EDM notch in Component specimen, figure from Smith et al. (2009).

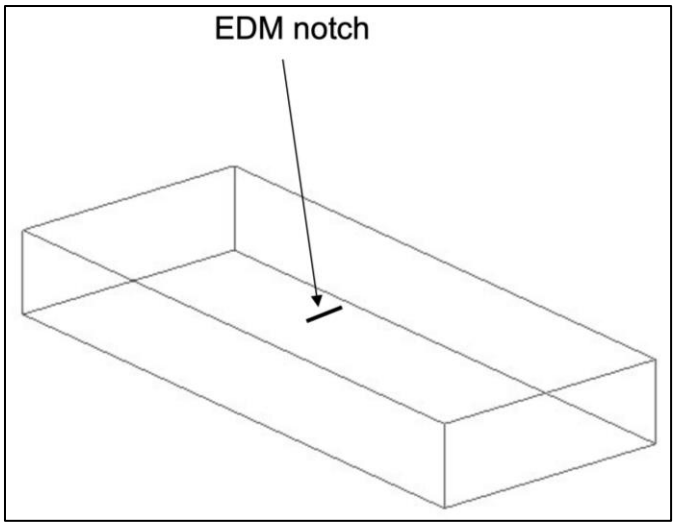


Figure C.2. Example 1 - EDM notch in Simple flat plate specimen, figure from Smith et al.

(2009).

Table C.1. Example 1 - Transfer Function Evaluation Inspection Measurements

Configuration	Flaw Size	Signal
Notches in Flat Plates	0.050	0.92
Notches in Flat Plates	0.107	2.60
Notches in Flat Plates	0.240	7.01
Notches in Flat Plates	0.259	4.42
Notches in Flat Plates	0.316	6.64
Notches in Flat Plates	0.372	9.12
Notches in Flat Plates	0.486	19.16
Notches in Flat Plates	0.524	10.55
Notches in Flat Plates	0.543	18.48
Notches in Flat Plates	0.600	14.89
Notches in Component	0.050	0.38
Notches in Component	0.107	1.09
Notches in Component	0.126	3.99
Notches in Component	0.202	4.40
Notches in Component	0.240	8.06
Notches in Component	0.372	4.22
Notches in Component	0.391	7.24
Notches in Component	0.410	14.37
Notches in Component	0.467	8.53
Notches in Component	0.562	8.95

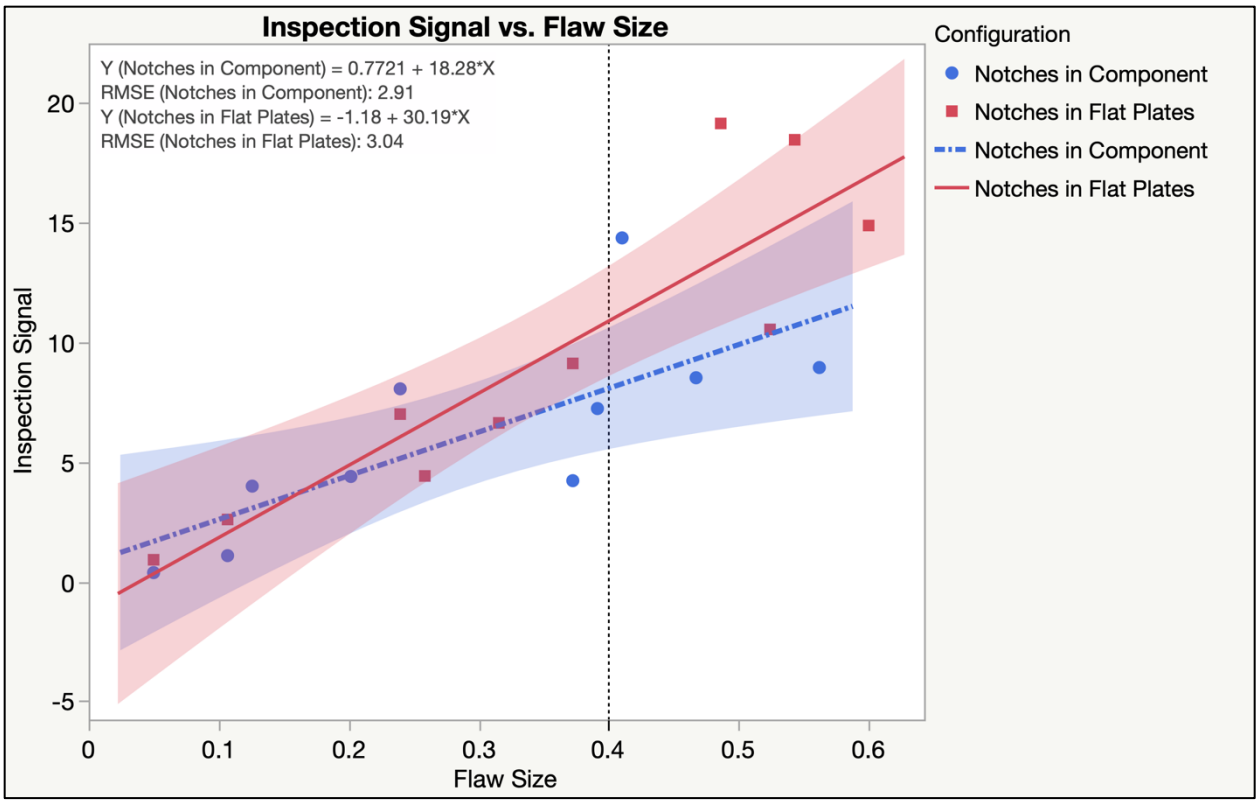


Figure C.3. Example 1 - inspection signal versus flaw size.

By inspection of Figure C.3, it appears that the model of the signal response for the Component specimen indicates a lower signal for the same size flaw compared to the Simple specimen at flaw sizes greater than about 0.2. In addition, the slope coefficients of the linear model are 18.3 and 30.2 for the Component and the Simple specimens respectively, which appear to be measurably different. However, from the plot it is apparent that the confidence intervals of the fitted linear models overlap at the CIFS of interest of 0.4, and therefore a difference in the average signal would not be detected in an assessment of similiarity, as described in Section 6.1. The upper and lower confidence intervals for the predicted signal at a flaw size of 0.4 are computed according to Section 5.2, as [5.5, 10.6] and [8.6, 13.2] for the Component and Simple specimen models, respectively. Comparing the upper bound from the Component model (i.e., 10.6) to the lower bound from the Simple model (i.e., 8.6) at a flaw size of 0.4, shows that the intervals overlap since 10.6 is greater than 8.6, which agrees with the visual interpretation. Therefore, a transfer function of the average signal level would not be recommended because it cannot be reliably detected in the presence of the inspection variability.

The residual error (i.e., RMSE) estimates in Figure C.3 are 2.9 and 3.0 for the Component and Simple specimens, respectively. Comparing the magnitude of the values directly without considering the variability may lead to a conclusion that the variability in the Component inspections is less than that of the Simple specimen. However, to determine if a difference can be reliably detected, the squared ratio of the standard deviations is computed as described in Section 6.2 as:

$$\left(\frac{s_{NC}}{s_{NS}}\right)^2 = \left(\frac{3.0}{2.9}\right)^2 = 1.1 \qquad \text{(C-1)}$$

The resulting value is compared to the cut-off value of 3.4 from the Table in Appendix B, using $n_c - 2 = 8$ and $n_s - 2 = 8$, since there are 10 inspection measurements on each specimen type. The NDE engineer would find that there is not a detectable difference in the variaiblity since 1.1 is less than 3.4. This result agrees with a visual interpretation of Figure C.3, which was computed as a demonstration of these calculations.

In this example, the NDE engineer would conclude that a difference in the average signal cannot be reliably detected nor a difference in the variability. Therefore, the NDE engineer would have demonstrated that similarity could be claimed between specimen types and a transfer function is not warranted. The practicial consequence of finding similarity is that a simulated-induced (e.g., fatigue crack) flaw size of 0.4, which is the same as the CIFS of interest in the flight component, could be used in a Simple specimen for POD demonstration. Alternatively, similarlity would support the use of a Standard NDE flaw size from 5009B, if the inspection method, instrument, and standardization were determined to be similar to the conditions to develop and/or confirm the Standard NDE flaw size.

Example 2 – Developing an Average Signal Transfer Function while Finding Similar Variability

This example is based on inspections of curved components compared to flat specimens with EDM notches as the fabricated flaws. The appropriate flaw size parameter was defined as the product of the length and the depth, denoted as LxD, and the flaw size and signal are log transformed to support a linear model of signal versus flaw size. The data are provided in Tables C.2 and C.3. There are 13 curved specimens and 39 flat specimens. A Forward transfer is desired where the CIFS of interest in the flight component is –9.7, and a simulated-induced flaw size in a Simple specimen will be specified for POD demonstartion. The data are plotted in Figure C.4, and the criticial initial flaw size of ln(LxD) = –9.7 is shown with a dashed vertical line. The linear RMSE values are shown in the upper left corner of the plot, and the shaded regions are 95% confidence intervals on the regression line.

Table C.2. Example 2 – EDM in Curved Inspection Data

Configuration	ln(LxD)	ln(Signal)
EDM in Curved	-9.87	2.28
EDM in Curved	-8.87	3.25
EDM in Curved	-9.81	2.41
EDM in Curved	-10.80	1.35
EDM in Curved	-9.76	2.15
EDM in Curved	-10.88	1.86
EDM in Curved	-9.84	1.98
EDM in Curved	-9.67	1.64
EDM in Curved	-10.07	1.89
EDM in Curved	-10.04	1.87
EDM in Curved	-10.96	1.40
EDM in Curved	-9.01	2.79
EDM in Curved	-9.92	2.27

Table C.3. Example 2 – EDM in Curved Inspection Data

Configuration	ln(LxD)	ln(Signal)
EDM in Flat	-11.74	1.03
EDM in Flat	-10.82	1.90
EDM in Flat	-10.01	2.66
EDM in Flat	-8.99	3.40
EDM in Flat	-10.82	1.47
EDM in Flat	-9.23	3.67
EDM in Flat	-9.90	3.25
EDM in Flat	-11.74	0.97
EDM in Flat	-10.01	2.14
EDM in Flat	-8.31	4.16
EDM in Flat	-10.01	2.31
EDM in Flat	-9.23	2.86
EDM in Flat	-8.99	3.19
EDM in Flat	-8.31	4.57
EDM in Flat	-9.90	2.42
EDM in Flat	-11.74	1.10
EDM in Flat	-9.23	3.69
EDM in Flat	-8.99	3.58
EDM in Flat	-8.99	3.45
EDM in Flat	-10.01	3.15
EDM in Flat	-10.93	1.59
EDM in Flat	-9.90	1.85
EDM in Flat	-9.90	2.91
EDM in Flat	-9.90	2.88
EDM in Flat	-8.31	3.76
EDM in Flat	-10.01	2.39
EDM in Flat	-10.82	1.59
EDM in Flat	-10.01	2.46
EDM in Flat	-10.93	1.18
EDM in Flat	-8.99	3.32
EDM in Flat	-8.99	3.57
EDM in Flat	-10.93	1.84
EDM in Flat	-11.74	1.09
EDM in Flat	-10.93	1.43
EDM in Flat	-8.31	3.83
EDM in Flat	-10.01	2.74
EDM in Flat	-10.01	1.93
EDM in Flat	-9.90	2.64
EDM in Flat	-10.01	2.88

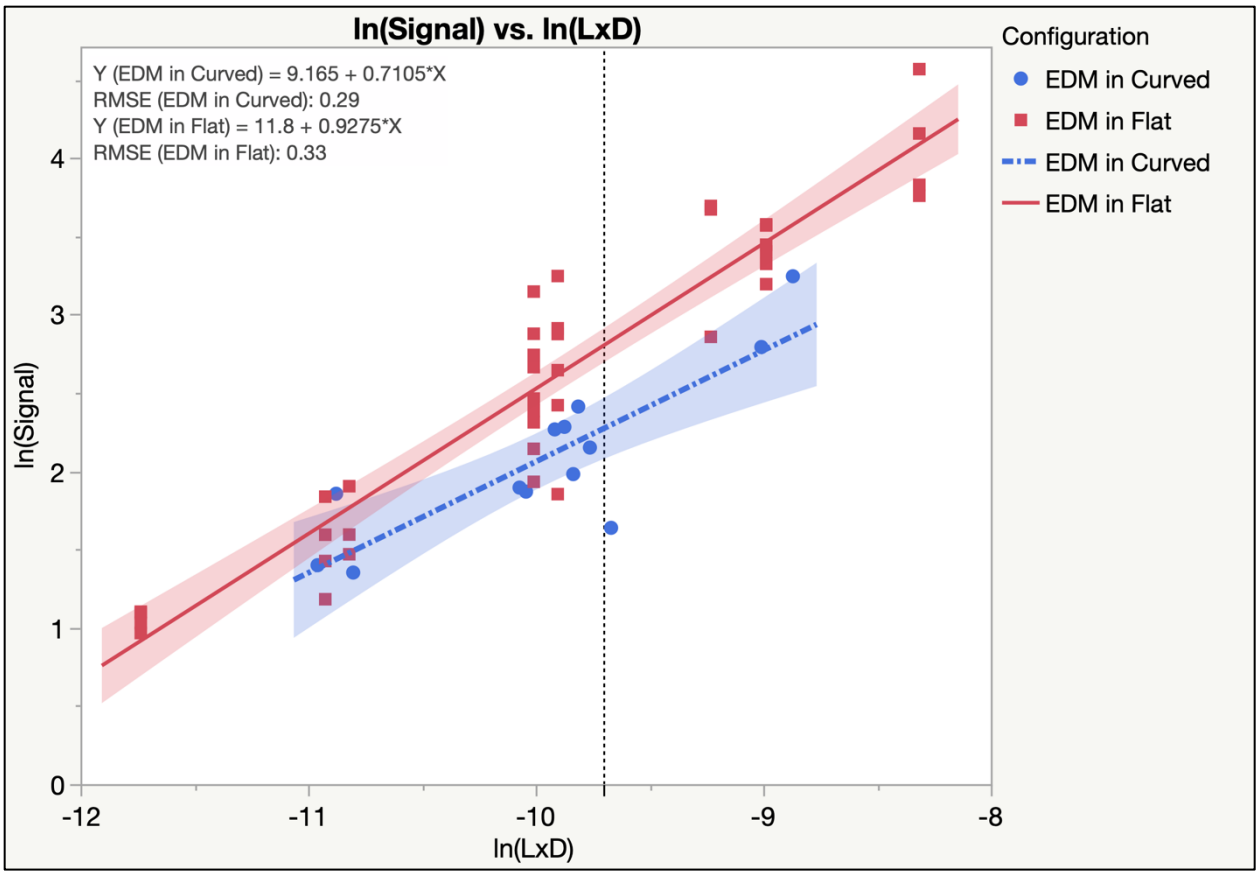


Figure C.4. Example 2 – ln(Signal) versus ln(LxD) plot with linear model coefficients

By way of general observations of Figure C.4, it shows a wider range of flaw sizes were inspected for the flat specimens compared to curved specimens. However, the CIFS is covered by both the specimens types, which ensures that there is no extrapolation involved. For the flat specimen, replicates of some flaw sizes are indicated by vertically stacked markers. Reflective of flight hardware inspection data, there is a single curved inspection data point noticeably away from the fitted line, which was not excluded from the analysis. This demonstrates a feature of this transfer function approach of fitting a model to a collection of inspection measurements rather than a point-to-point comparison which may be negatively affected by this potentially anomolous data point. Of course, if there is an explainable cause for removal of this data point to be removed, that is acceptable. Finally, the different specimen-type sample sizes are accommodated by modeling the configurations individually, and the benefit of the larger sample size and wider range of flaw sizes for the flat specimens results in its narrower confidence interval.

By observation, the confidence intervals at a flaw size of –9.7 do not overlap, and therefore similarity is not declared and a transfer function will be developed for the average flaw size. Using the formulae given in Section 6.1, the lower and upper bounds of the confidence intervals at the flaw size of –9.7 are [2.08, 2.47] and [2.70, 2.91] for the curved and flat specimens, respectively. Comparing the lower confidence bound of the flat specimen to the upper confidence bound of the curved specimen shows that 2.70 is greater than 2.47, which is consistent with our graphical finding of not declaring similiarity.

Based on the residual errors (i.e., RMSE) of 0.29 and 0.33 for the curved and flat specimens, there does not appear to be practical difference. The squared ratio of the standard deviations is computed as described in Section 6.2 as:

$$\left(\frac{s_{NC}}{s_{NS}}\right)^2 = \left(\frac{0.29}{0.33}\right)^2 = 0.77 \tag{C-2}$$

The resulting value can be compared to the cut-off value of 2.3 from the Table in Appendix B, using $n_c - 2 = 11$ and $n_s - 2 = 37$, since there are 13 inspection measurements of the curved specimens and 39 measurements of the flat specimen. The Appendix B table provides $n_s - 2$ to 20. However, the maximum value of 2.3 could be used in this case as an approximation, since the cut-off values are increasing slowly for larger sample sizes in the table. The exact cut-off value for this case can be computed as 2.1 from the distribution shown in Section 6.2. Therefore, a detectable difference in the variaiblity is not found, since 0.78 is less than 2.3.

As a note, it might be misleading to compare the width of the confidence intervals (i.e., shaded regions) in Figure C.4 in determining if there is a difference in variability rather than the residual errors (i.e., RMSE). The confidence intervals include the residual error in their calculations, as shown in Section 5.2. However, they include a measure of the spread of the flaw sizes, S_{xx}, and in this case the wider spread (i.e., range of flaw sizes) of flat specimens strongly contributes to the narrower confidence intervals for the flat specimens, even though the residual errors are about the same. The residual errors are the difference between each inspection data point and their respective fitted line, and it is evident that the spread about each line is about the same (i.e., ignoring the shaded confidence regions), which results in a similar RMSE.

Based on the similarity evaluation, a Forward transfer function for the average signal will be developed. However, there will not be a transfer based on a difference in variability. Following the process in Section 7.1.1, the transferred flaw size from Component specimen to Simple specimen provides a flaw size in the Simple specimen with an equivalent signal level as the CIFS in the Component specimen, computed as:

$$\hat{d}_{mean} = \frac{\left(\hat{\beta}_{0_{NC}} + \hat{\beta}_{1_{NC}} x_t\right) - \hat{\beta}_{0_{NS}}}{\hat{\beta}_{1_{NS}}} \tag{C-3}$$

$$\hat{d}_{mean} = \frac{(9.165 + 0.7105(-9.7)) - 11.8}{0.9275} = -10.27 \tag{C-4}$$

Therefore, a simulated-induced flaw (e.g., fatigue crack) that has a ln(LxD) of −10.27 could be used in a Simple specimen for POD demonstration. Figure C.5 includes the transferred flaw size of −10.27 to illustrate the results of the formula used to compute it.

The complete numerical calculations for this example are provided in Table C.4 using a modified version of the Excel® spreadsheet to accommodate the sample sizes.

Table C.4. Numerical Calculations for Example 2 using a Modified Version of the Excel® Template

FORWARD CASE - Specimen Type Similarity and Transfer Function Calculations Input Output

Reference NASA TM "Guidebook for Assessing Similarity and Implementing Empirical Transfer Functions for Probability of Detection (POD) Demonstrations for Single-Hit Nondestructive Evaluation (NDE) Methods"

Scenario: Given a target naturally occurring flaw size to be detected in a flight component, typically from fracture analysis
What size simulated-induced flaw should be specified in a simple specimen for each inspector to demonstrate POD?

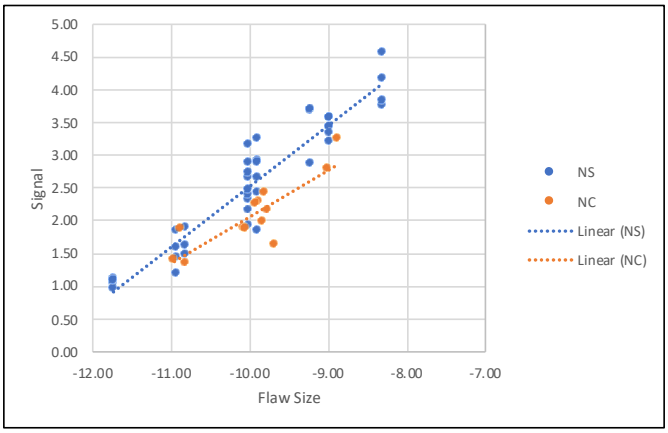
Summary of Analysis: Input Target Flaw Size, and Resultant POD Demonstration Flaw Size

Target Flaw Size	-9.70	Naturally occuring flaw size in flight component
POD Demo Flaw Size	-10.27	Simulated-Induced flaw size in a simple specimen

Transfer Evaluation Measurements used to Assess Similatiry and Specimen Transfer Function(s)

Simulated-Fabricated Flaws in Simple

Flaw Size (x)	Signal (y)
-11.74	1.03
-10.82	1.90
-10.01	2.66
-8.99	3.40
-10.82	1.47
-9.23	3.67
-9.90	3.25
-11.74	0.97
-10.01	2.14
-8.31	4.16
-10.01	2.31
-9.23	2.86
-8.99	3.19
-8.31	4.57
-9.90	2.42
-11.74	1.10
-9.23	3.69
-8.99	3.58
-8.99	3.45
-10.01	3.15
-10.93	1.59
-9.90	1.85
-9.90	2.91
-9.90	2.88
-8.31	3.76
-10.01	2.39
-10.82	1.59
-10.01	2.46
-10.93	1.18
-8.99	3.32
-8.99	3.57
-10.93	1.84
-11.74	1.09
-10.93	1.43
-8.31	3.83
-10.01	2.74
-10.01	1.93
-9.90	2.64
-10.01	2.88



Simulated-Fabricated Flaws in Component

Flaw Size (x)	Signal (y)
-9.87	2.28
-8.87	3.25
-9.81	2.41
-10.80	1.35
-9.76	2.15
-10.88	1.86
-9.84	1.98
-9.67	1.64
-10.07	1.89
-10.04	1.87
-10.96	1.40
-9.01	2.79
-9.92	2.27

Modeling of Transfer Evaluation Inspection Measurements

Estimate the Coefficients of Linear Models

Signal = $\beta 0 + \beta 1 * \text{Flaw} + \varepsilon$

Model Coefficients for Simple		Model Coefficients for Component	
$\beta 1$_hat_NS	0.9275	$\beta 1$_hat_NC	0.7105
$\beta 0$_hat_NS	11.80	$\beta 0$_hat_NC	9.17
s_NS	0.33	s_NC	0.29

Testing Similarity of Average Signal

Can we reliably detect a difference in the average signal for the same size Flaw in the Simple specimen compared to the Component?
If the confidence intervals do not overlap at the target flaw size, then we claim to detect an effect on the average signal response
ASSUME: Signal for Component is less than Signal for Simple specimen at Target Flaw Size

95% Confidence Interval of the Fitted Line at Target Flaw Size

Confidence Interval for Simple			Confidence Interval for Component		
y_hat_NS	2.80	at Target Flaw Size	y_hat_NC	2.27	at Target Flaw Size
t-value for NS	2.03	95% CI	t-value for NC	2.20	95% CI
delta_y_CI_NS	0.11		delta_y_CI_NC	0.20	
Upper CI_NS	2.91		Upper CI_NC	2.48	
Lower CI_NS	2.70		Lower CI_NC	2.07	

Is the Lower CI in Simple > Upper CI in Component? TRUE

If "TRUE" then we can detect a difference in the signal level at the target flaw size

Testing Similarity in Signal Variability

Can we reliably detect a difference in the signal variability between the Simple and Component specimens?
Based on the squared ratio of the RMSE from the regression models

(s_NC^2)/(s_NS^2)	0.81	
Cut-off value to Exceed	2.1	Finv, 95% one-sided test of Comp_var > Simple_var

Is the ratio of the squared RMSEs greater than the Cut-off Value? FALSE

If "TRUE" then we can detect a difference in the signal variability between the Simple and Component

Average Signal Transfer Function

Only if a difference in the average signal is detected
Adjust the average signal from the Component to the Simple Specimen

Specimen Transfer of Flaw Size from Component to Simple Specimen

dhat_mean	**-10.27**	flaw size units

Signal Variability Transfer Function

Only if a difference in the the signal variability is detected
Adjust the average signal due to additional variability in the Component compared to the Simple Specimen

Convert the signal variaiblity to flaw size variability using estimated slope coefficients

Fabricated Flaw in Component	0.41	flaw size units
Fabricated Flaw in Simple	0.35	flaw size units
Δdhat_var	**0.00**	Adjustment to Flaw Size Units (Difference)

Combining the Average Signal and Variaibility Specimen Type Transfers

dhat = dhat_mean - Δdhat_var

dhat	**-10.27**	Simulated-Induced flaw size in Simple specimen for POD Demo

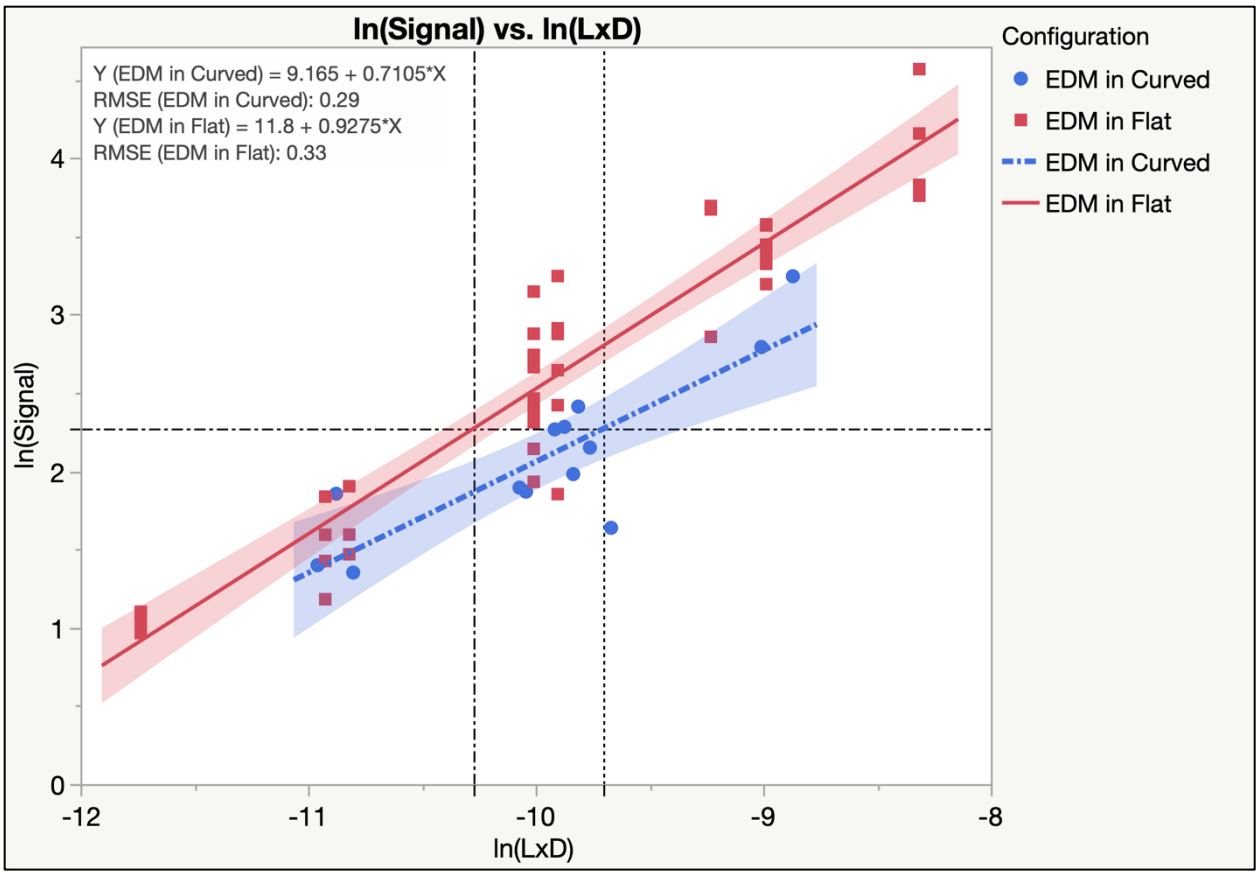


Figure C.5. Example 2 – ln(signal) versus ln(LxD) with transferred flaw size reference lines.

References

Meeker, W.Q. (2021): Informal communication, Sept. 13th, 2021.

Smith, K.; and Thompson, B., Brasche, L. (2009): "Model-based POD Successes and Opportunities," 2009 MAPOD Working Group

Appendix D. Evaluating Similarity and Transfer Function Methodology for Flaw-Type Transfers

The specimen-type transfer function presented in this guidebook enables the specification of a representative induced flaw size (e.g., fatigue cracks) for POD demonstration, and it is the recommended approach since it replicates flaw-to-flaw variability of naturally occurring flaws. However, flaw-type transfers have been used previously for some limited applications on NASA programs with FCB approval. Flaw type transfer functions estimate flaw sizes between two types of flaws (e.g., induced and fabricated) that produce similar average signal responses. Using this flaw type transfer function, fabricated flaws are sized for POD demonstration tests in the hardware component to be inspected. As previously noted, there does not exist documented methodology for applying this type of transfer function for NASA programs and there is a concern that in some cases the approaches used may not address the increase in variability of signal response for induced and naturally occuring flaws as compared to fabricated flaws used in the POD test. While the specimen type transfer function developed in this guidebook is recommended as the preferred method, it is noted that the same statistical approaches can be applied to flaw type transfer functions. As such, guidance for flaw type transfer functions is provided that includes steps to incorporate the difference in the mean signal response, and the variability in signal responses between flaw types.

In Figure D.1, a flaw-type transfer occurs across the bottom row of the diagram going from an induced flaw size to fabricated flaw size, where both flaw types are in a Simple specimen. In practice, a common flaw type transfer is between fatigue cracks (i.e., induced flaws) and EDM notches (i.e., fabricated flaws), and the transfer function is used to specify a representative notch size placed in a Component specimen that can be used to optimize the NDE technique, study the effects of varying geometry and access restrictions across the component, and/or assess detection capability. When transferring between flaw types in Simple specimens, it is assumed that the flaw-type transfer is independent of the specimen type, and therefore, the transfer function can be used to transfer naturally occurring flaws in the flight component to fabricated flaws in a Component specimen. For a flaw-type transfer, a naturally occurring flaw in a component of interest is specified, typically from a fracture analysis, and the size of an equivalent fabricated flaw in a Component specimen is sought, see Figure D.1.

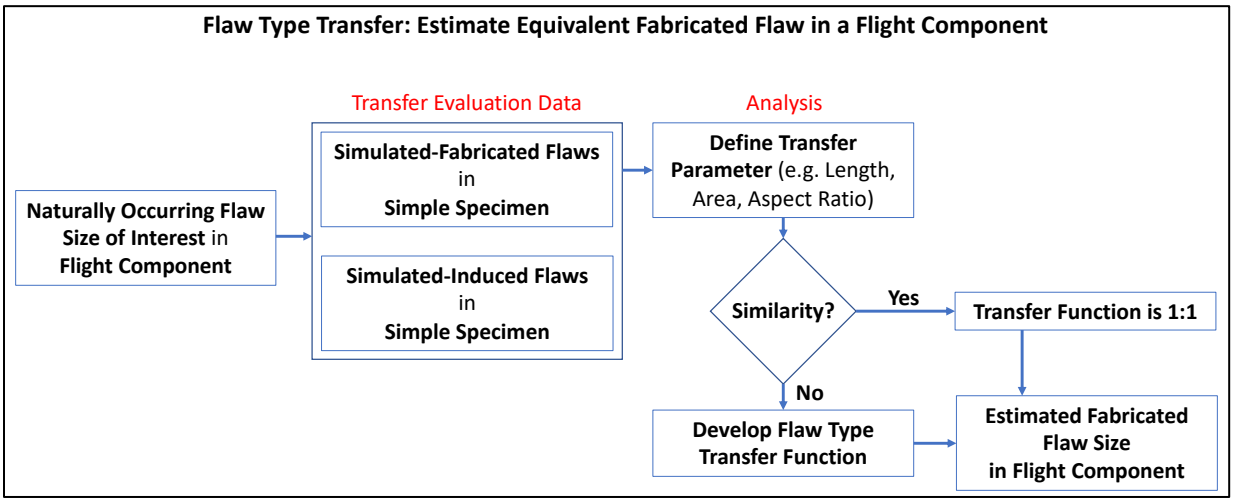


Figure D.1. Flaw-type transfer function block diagram.

Consistent with the specimen-type transfer presented in this guidebook, a similarity assessment is conducted to determine if a flaw-type difference in the signal response can be reliably detected in the presence of the inspection variability before a transfer function is developed. If a difference is not detected, then similarity could be declared (i.e., a 1-1 transfer function). In this case, the fabricated flaw size is the same as the naturally occurring flaw size in terms of signal response. If a difference in the average signal level and/or the signal variability is detected, then a transfer function could be pursued to adjust the fabricated flaw size. Similar to the specimen-type, this flaw-type transfer function approach is a point-type transfer since it operates on a single CIFS of interest not the full POD relationship across a range of flaw sizes.

Transfer evaluation data are collected, as shown in Table C.1, from induced flaws and fabricated flaws in Simple specimens, and all measurements are performed by a single inspector using the same NDE method and standardization procedure. For the fabricated flaws in a Simple specimen, three unique flaw sizes covering the flaw size of interest are recommended with two replicates of each size, as a minimum data set. For induced flaws in a Simple specimen, it is recommended to have five unique flaw sizes with five replicates of each nominal size, as a minimum data set. The additional flaw sizes and replicates are due to the expectation of greater flaw-to-flaw variability of induced flaws compared to fabricated flaws. It is recognized that replicated induced flaw sizes will not be identical size due to a lack of precise production. However, the replicates are intended to provide a cluster of inspections around the same nominal flaw size to characterize flaw-to-flaw variability independent of the ability to model the signal relationship across flaw sizes. It is recommended that the induced flaws in the specimens are independently characterized using another NDE technique that enables sizing of the flaw or destructively to reveal the induced flaw morphology. A discussion and guidance on flaw size selection is provided in Section 5.1.

The analysis of the transfer evaluation data begins with estimating a regression model of the signal to flaw size relationship, evaluating similarity of the average signal and its variability, and developing transfer functions, if deemed appropriate. Section 5.2 discusses modeling considerations. The Forward case of the specimen-type transfer function is generally applicable by replacing references to fabricated flaws in a Component specimen with induced in a Simple specimen, and the subscript CS is introduced to denote induced flaws in a Simple specimen. This substitution of variables is provided in a condensed summary in the following sections and further details can be found in the main body of this guidebook.

Modeling the Transfer Function Evaluation Measurements

A simple linear regression model in the form of $y = \beta_0 + \beta_1 x + \varepsilon$ is fit from the measurements or their transformed variables obtained on each flaw type, where β_0 is the intercept, β_1 is the slope, and ε is the random error that is assumed to be an independently, identically distributed, normal random variable with a mean of 0 and a variance of σ^2. From the linear regression, the model parameters are estimated as $\hat{\beta}_0$, $\hat{\beta}_1$, and s, which is the estimated standard deviation of ε for each flaw type. See Montgomery et al. (2012) for a more complete description of linear regression.

The estimated models for each flaw type are shown in Equations D-1 and D-2, and these expressions enable the prediction of inspection measurement signals, y, from each flaw type as a function of flaw size, x. The circumflex notation indicates an estimated parameter or quantity.

$$\hat{y}_{NS} = \hat{\beta}_{0_{NS}} + \hat{\beta}_{1_{NS}} x_{NS} \tag{D-1}$$

$$\hat{y}_{CS} = \hat{\beta}_{0_{CS}} + \hat{\beta}_{1_{CS}} x_{CS} \tag{D-2}$$

The statistical uncertainty in the model for each flaw type can be expressed with a confidence interval around the fitted linear model. The confidence interval is based on $\hat{y} \pm \Delta\hat{\mu}_{CI}$, where the formulation for $\Delta\hat{\mu}_{CI}$ is provided in Equation D-3.

$$\Delta\hat{\mu}_{CI} = t_{(\alpha/2,n-2)}\, s\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}} \tag{D-3}$$

In Equation D-3, n is the number of measurements in the dataset used to estimate the model coefficients, $t_{(\alpha/2,n-2)}$ is the percentile from a student's t-distribution for 100(1- α)% confidence interval with n-2 degrees of freedom, s is the estimated standard deviation of the signal, x_0 is a specific criticial initial flaw size, and $\bar{x}$ is the average flaw size in the dataset. It is recommended to use $\alpha = 0.05$ providing a 95% confidence interval, and *t*-values are provided in Appendix A for datasets containing $n - 2 \leq 20$ inspection measurements. The term S_{xx} is the sum of squared difference between each individual flaw size in the transfer evaluation dataset and the average flaw size in the dataset, computed as:

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{D-4}$$

Evaluating Similarity Between Induced and Fabricated Flaws

After estimating the linear models and confidence intervals for both flaw types, statistical tests are performed to determine if a difference can reliably be detected between the average signal level and signal variability from fabricated and induced flaws. If no differences are reliably detected, then similarity is supported as discussed in NASA 5009B. In this case, the transfer function is 1-1, and the naturally occurring CIFS desired to be detected in the flight component is considered equal to the fabricated flaw.

Detecting a Difference in Average Inspection Signal Between Flaw Types

The confidence intervals around the fitted lines are used to detect whether there is a difference in the average signal for the same size flaw in both flaw types. If the confidence intervals overlap, then no difference can be reliably detected at that flaw size. In addition, if the confidence intervals do not overlap, then a difference in the average signal is detected.

Detecting a Difference in the Signal Variability Between Flaw Types

The estimated signal variability from each flaw type is used to determine if a difference in the signal variability between specimen types can be reliably detected. The RMSE from the regression represents the average variability of individual inspection measurements across the range of flaw sizes in the transfer function evaluation measurement datasets, and therefore it is used in the following analyses as being independent of flaw size.

A hypothesis testing approach is employed to determine if a difference in the signal variability can be reliably detected, where the null (H_0) and alternative (H_1) hypotheses are based on ratios of the variances from each flaw type as:

$$H_0: \frac{\sigma_{CS}^2}{\sigma_{NS}^2} = 1 \tag{D-5}$$

$$H_1: \frac{\sigma_{CS}^2}{\sigma_{NS}^2} > 1 \tag{D-6}$$

where σ^2 represents the true signal variability. To conduct the test, an observed test statistic (f_{obs}) is computed from the regression RMSE standard deviations from the transfer function evaluation measurements as:

$$f_{obs} = \left(\frac{s_{CS}}{s_{NS}}\right)^2 \tag{D-7}$$

If the ratio of the estimated signal variabilities is close to 1, then a difference in variability is not detected. As the ratio increases, the evidence becomes stronger that there is a detectable difference. The test statistic is compared to a value from an *F*-distribution, $F_{(\alpha,n_{CS}-2,n_{NS}-2)}$, which is a percentile value associated with 100(1- α)% probability with $(n_{CS}-2)$ and $(n_{NS}-2)$ degrees of freedom. This *F*-value is known as a cut-off value for the hypothesis test. In this expression, n_{CS} and n_{NS} are the number of inspection measurements from the induced and fabricated flaws, respectively. It is recommended to use $\alpha = 0.05$, and *F*-values are provided in Appendix B for datasets containing $(n-2) \leq 20$ inspection measurements. If the test statistic exceeds the cut-off value from the *F*-distribution, such that:

$$f_{obs} > F_{(\alpha,n_{CS}-2,n_{NS}-2)} \tag{D-8}$$

then a difference in variability between the flaw types is detected. This formulation assumes that the variability of the induced flaws will typically be larger than the variability of the fabricated flaws.

Estimating Transfer Functions when Similarity is not declared

If a difference in the average signal level and/or the variability is detected between flaw types, then a transfer function approach could be developed. In practice, a naturally occurring CIFS (x_t) in a flight component is provided and a representative fabricated flaw size $\hat{d}$ is sought to be specified.

Transferring the Average Signal Level

To transfer the average signal level from induced to fabricated flaws, x_{CS} is replaced with x_t in Equation D-2 and the signal level ($\hat{y}_t$) from an induced flaw of size x_{CS} is predicted as:

$$\hat{y}_t = \hat{\beta}_{0_{CS}} + \hat{\beta}_{1_{CS}} x_t \tag{D-9}$$

Then, x_{NS} is replaced by $\hat{d}_{mean}$, which is the fabricated flaw size, and $\hat{y}_{NS}$ is replaced by $\hat{y}_t$.

$$\hat{y}_t = \hat{\beta}_{0_{NS}} + \hat{\beta}_{1_{NS}} \hat{d}_{mean} \tag{D-10}$$

Setting these two equations equal to each other and rearranging to estimate the fabricated flaw size $\hat{d}_{mean}$, provides the expression:

$$\hat{d}_{mean} = \frac{\left(\hat{\beta}_{0_{CS}} + \hat{\beta}_{1_{CS}} x_t\right) - \hat{\beta}_{0_{NS}}}{\hat{\beta}_{1_{NS}}} \tag{D-11}$$

This equation estimates the fabricated flaw size using the coefficients of the linear models from both flaw types and the specified CIFS to be detected in the flight component.

Transferring the Signal Variability

If a difference in the signal variability is detected between the flaw types, then the average signal level ($\hat{d}_{mean}$) is adjusted by $\Delta\hat{d}_{var}$. If no difference in the average signal level was detected, where

$\hat{d}_{mean} = x_t$, then an adjustment due to variability can be considered. The procedure converts the standard deviation of the signal variability (s) to an adjustment in flaw size ($\Delta\hat{x}_{var}$) using the estimated slopes from the linear models for each flaw type as:

$$\Delta\hat{x}_{var_{NS}} = \frac{s_{NS}}{\hat{\beta}_{1_{NS}}} \tag{D-12}$$

$$\Delta\hat{x}_{var_{CS}} = \frac{s_{CS}}{\hat{\beta}_{1_{CS}}} \tag{D-13}$$

Assuming that the baseline signal variability from the fabricated flaws is included in the signal variability of the induced flaws, the adjustment is based on the increase in variability in the induced flaws as:

$$\Delta\hat{d}_{var} = \Delta\hat{x}_{var_{CS}} - \Delta\hat{x}_{var_{NS}} \tag{D-14}$$

Reporting the Transferred Fabricated Flaw Size

Combining the transfer of the average signal and/or adjusting for increased signal variability, the representative fabricated flaw size is computed as:

$$\hat{d} = \hat{d}_{mean} - \hat{d}_{var} \tag{D-15}$$

Flaw-Type Transfer Function Case Study Example

In an investigation by Kane and Koshti (2008), eddy current inspections were performed with induced flaws (e.g., fatigue cracks) and fabricated flaws (e.g., EDM notches) in aluminum specimens. The length times depth of the flaw (LxD) is used to model the signal versus flaw size relationship, where LxD is an approximate measure of flaw surface area. A naturally occurring flaw with LxD of 1.5E-03 is specified by fracture analysis and a representative fabricated flaw size is sought. The data are provided in Table D.1 and plotted in Figure D.2. There are six EDM notches and 25 fatigue cracks specimens. In Figure D.2, the CIFS is indicated with a dashed vertical line. The linear regression model coefficients and residual errors (i.e., RMSE) are shown in the upper left corner of the plot, and the shaded regions are 95% confidence intervals on the regression lines.

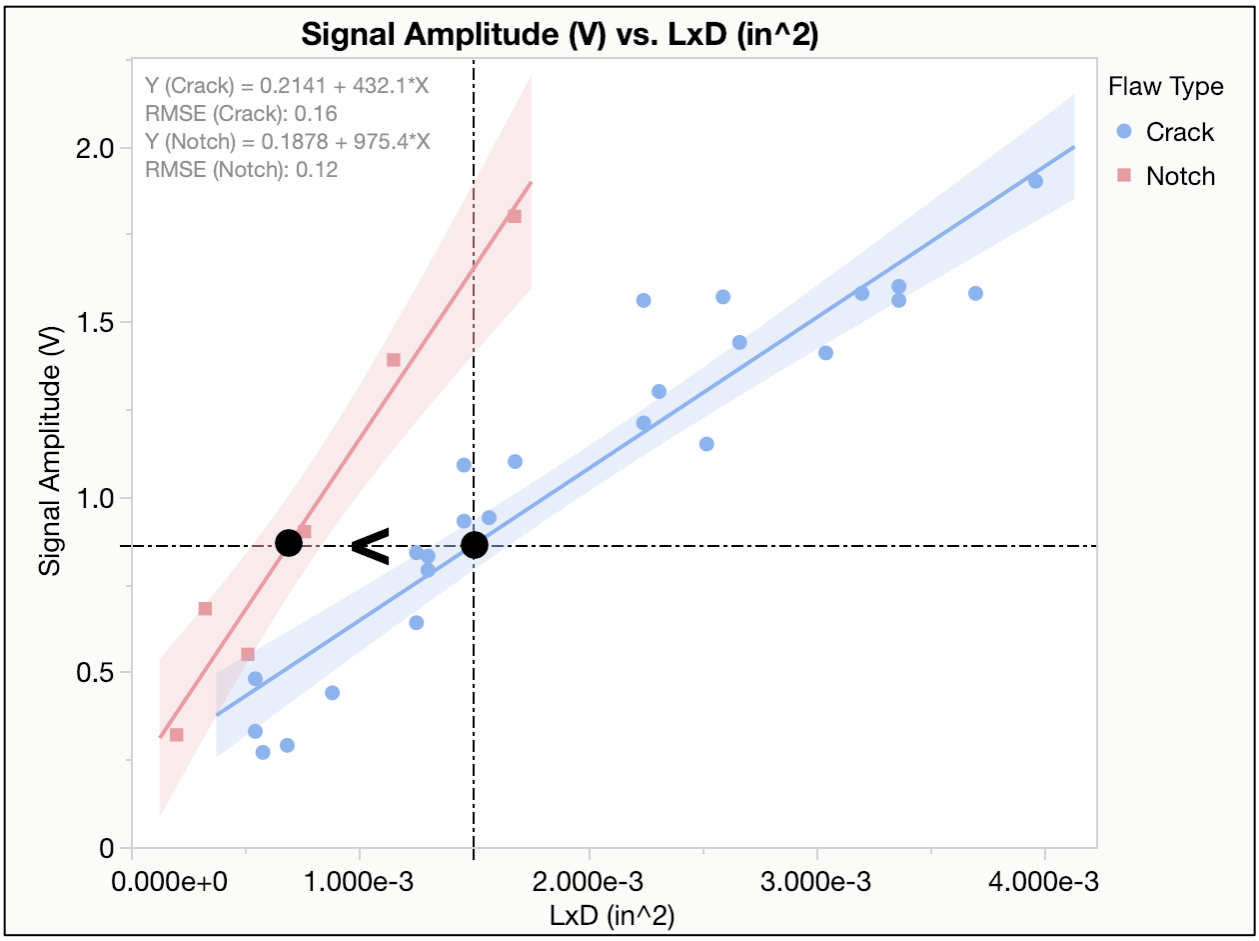


Figure D.2. Flaw type transfer case study example plot of signal versus flaw size.

Table D.1. Flaw-type Transfer Case Study Data

Flaw Type	LxD (in^2)	Signal Amplitude (V)
Crack	1.3005E-03	0.83
Crack	3.0420E-03	1.41
Crack	3.9605E-03	1.90
Crack	2.2445E-03	1.21
Crack	5.7800E-04	0.27
Crack	1.4580E-03	0.93
Crack	3.2000E-03	1.58
Crack	2.2445E-03	1.56
Crack	1.4580E-03	1.09
Crack	1.5680E-03	0.94
Crack	2.3120E-03	1.30
Crack	5.4450E-04	0.48
Crack	5.4450E-04	0.33
Crack	1.2500E-03	0.64
Crack	2.5920E-03	1.57
Crack	3.3620E-03	1.60
Crack	1.2500E-03	0.84
Crack	6.8450E-04	0.29
Crack	2.5205E-03	1.15
Crack	3.6980E-03	1.58
Crack	2.6645E-03	1.44
Crack	1.6820E-03	1.10
Crack	3.3620E-03	1.56
Crack	8.8200E-04	0.44
Crack	1.3005E-03	0.79
Notch	2.0000E-04	0.32
Notch	3.2500E-04	0.68
Notch	5.1200E-04	0.55
Notch	7.6000E-04	0.90
Notch	1.1500E-03	1.39
Notch	1.6800E-03	1.80

In Figure D.2, there is a wider range of crack flaw sizes compared to EDM notches. However, the CIFS is covered by both flaw types, thereby avoiding extrapolation. By observation, the confidence intervals at a flaw size of 1.5E-03 do not overlap, and (i.e., similarity is not declared), and a transfer function will be developed for the average signal. Using the formulae given in Section 5.2, the lower and upper bounds of the confidence intervals at the flaw size of 1.5E-03 are [0.79, 0.93] and [1.41, 1.90] for the crack and notch flaw types, respectively. Comparing the lower confidence bound of the notch specimen to the upper confidence bound of the crack specimen shows that 1.90 is greater than 0.79, which is consitent with the visual inspection of the plot.

Based on the residaul errors (i.e., RMSE) of 0.16 and 0.12 for the crack and notch specimens, respectively, the squared ratio of the standard deviations is computed as:

$$\left(\frac{s_{CS}}{s_{NS}}\right)^2 = \left(\frac{0.16}{0.12}\right)^2 = 1.59 \quad \text{(D-16)}$$

The resulting value is compared to the cut-off value of 5.8 from Table D.1, substituing n_{CS} for n_{NC}, results in $n_{CS} - 2 = 23$ and $n_{NS} - 2 = 4$, since there are 25 cracks and six notches inspected. Appendix B provides $n_{CS} - 2$ to 20, however the maximum value of 5.8 could be used in this case as an approximation, since it is evident that the cut-off values are not increasing with larger n_{CS} sample sizes in the table across the row of $n_{NS} - 2 = 4$. The cut-off value for this case can be computed as 5.8 from the F-distribution. Therefore, a detectable difference in the variaiblity is not found, since 1.59 is less than 5.8.

Based on the similarity evaluation finding a difference in average signal, a transfer function for the average signal is developed, and there will not be a transfer based on a difference in variability since it was not detected. The transferred flaw size from induced to fabricated flaw size is computed as:

$$\hat{d}_{mean} = \frac{\left(\hat{\beta}_{0_{CS}} + \hat{\beta}_{1_{CS}} x_t\right) - \hat{\beta}_{0_{NS}}}{\hat{\beta}_{1_{NS}}} \quad \text{(D-17)}$$

$$\hat{d}_{mean} = \frac{(0.2141 + 432.1(1.5\text{E-}03)) - 0.1878}{975.4} = 6.9\text{E-}04 \quad \text{(D-18)}$$

Therefore, a fabricated flaw (e.g., EDM notch) that has a LxD of 6.9E-04 is considered representative of a naturally occurring flaw with a LxD of 1.5E-03.

It seems imperative to comment that this case study results in a 2.2 ratio of LxD parameter of flaw size, and if the flaw depth is the same in both flaw types, then this is equivalent to a 2.2 ratio of flaw length to depth. While this is consistent with a commonly used 'rule-of-thumb' 2-1 ratio of flaw length, it should be noted that crack depth was not independently characterized in the data provided by Kane and Koshti (2008), rather it was assumed that depth is equal to L/2, assuming a constant L/D = 2 ratio. This highlights the importance of independent characterization of the induced flaw sizes used in the transfer function development. The fabricated flaw depth was based on L/D = 2, which could be controlled in fabrication. In summary, based on the design of this study, signal response of LxD is not independent of length, and therefore the effect of length cannot be evaluated independently from LxD. This case study is provided an example of the methodology, and the specific numerial results should not be generalized.

Utilization of the representative fabricated flaw size

The utility of the representative fabricated flaw size in a flight component depends on the application. However, it is not preferred to use fabricated flaws in POD demonstration to satisfy 5009B Special NDE requirements, since it may not adequately represent flaw-to-flaw variability of naturally occurring flaws. It is envisioned that representative fabricated flaws can be used to optimize the NDE technique or study the effects of varying geometry and access restrictions across the component of interest.

In the case study example, there was not a detectable difference in flaw-to-flaw variability between flaw types using the method presented. Therefore, in this specific case, a Special NDE POD demonstration strategy may include fabricated flaws in addition to demonstrating POD capability with induced flaws that are expected to better represent the variation in flaw morphology of naturally occurring flaws. The details of the Special NDE demonstration strategy would be

developed by the NDE engineer cognizant of the NDE technique and the inspection application and subject to review by the FCB, and therefore, they are not specified in this guidebook.

Reference for Flaw Type Transfer Appendix:

Kane, M.; and Koshti, A. (2008): "Comparison of Signal Response Between EDM Notches and Cracks in Eddy-Current Testing," September 29, 2008, unpublished NASA internship report.

Montgomery; Peck; and Vining (2012): Introduction to Linear Regression Analysis, 5th Edition, John Wiley & Sons, Hoboken, 2012.

Appendix E. Forward Case: Excel® Spreadsheet Evaluating Similarity and Developing Transfer Functions

FORWARD CASE - Specimen Type Similarity and Transfer Function Calculations Input Output

Reference NASA TM "Guidebook for Assessing Similarity and Implementing Empirical Transfer Functions for Probability of Detection (POD) Demonstrations for Single-Hit Nondestructive Evaluation (NDE) Methods"

Scenario: Given a target naturally occurring flaw size to be detected in a flight component, typically from fracture analysis
What size simulated-induced flaw should be specified in a simple specimen for each inspector to demonstrate POD?

Summary of Analysis: Input Target Flaw Size, and Resultant POD Demonstration Flaw Size

Target Flaw Size	0.065	Naturally occuring flaw size in flight component
POD Demo Flaw Size	0.032	Simulated-Induced flaw size in a simple specimen

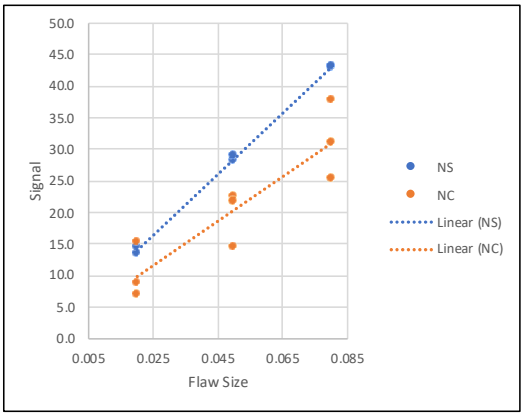
Transfer Evaluation Measurements used to Assess Similatiry and Specimen Transfer Function(s)

Simulated-Fabricated Flaws in Simple

Flaw Size (x)	Signal (y)
0.020	13.5
0.020	14.4
0.050	28.2
0.050	29.0
0.080	42.8
0.080	43.2

Simulated-Fabricated Flaws in Component

Flaw Size (x)	Signal (y)
0.020	8.8
0.020	15.1
0.020	7.0
0.050	22.5
0.050	21.7
0.050	14.5
0.080	31.1
0.080	37.8
0.080	25.4



Modeling of Transfer Evaluation Inspection Measurements

Estimate the Coefficients of Linear Models

Signal = ! 0 + ! 1*Flaw + "

Model Coefficients for Simple	
! 1_hat_NS	484
! 0_hat_NS	4.31
s_NS	0.45

Model Coefficients for Component	
! 1_hat_NC	352
! 0_hat_NC	2.82
s_NC	4.71

Testing Similarity of Average Signal

Can we reliably detect a difference in the average signal for the same size Flaw in the Simple specimen compared to the Component?
If the confidence intervals do not overlap at the target flaw size, then we claim to detect an effect on the average signal response
ASSUME: Signal for Component is less than Signal for Simple specimen at Target Flaw Size

95% Confidence Interval of the Fitted Line at Target Flaw Size

Confidence Interval for Simple			**Confidence Interval for Component**		
y_hat_NS	35.8	at Target Flaw Size	y_hat_NC	25.7	at Target Flaw Size
t-value for NS	2.78	95% CI	t-value for NC	2.36	95% CI
delta_y_CI_NS	0.60		delta_y_CI_NC	4.91	
Upper CI_NS	36.4		Upper CI_NC	30.6	
Lower CI_NS	35.2		Lower CI_NC	20.8	

Is the Lower CI in Simple > Upper CI in Component? **TRUE**

If "TRUE" then we can detect a difference in the signal level at the target flaw size

Testing Similarity in Signal Variability

Can we reliably detect a difference in the signal variability between the Simple and Component specimens?
Based on the squared ratio of the RMSE from the regression models

(s_NC^2)/(s_NS^2)	107.6	
Cut-off value to Exceed	6.1	Finv, 95% one-sided test of Comp_var > Simple_var

Is the ratio of the squared RMSEs greater than the Cut-off Value? **TRUE**

If "TRUE" then we can detect a difference in the signal variability between the Simple and Component

Average Signal Transfer Function

Only if a difference in the average signal is detected
Adjust the average signal from the Component to the Simple Specimen

Specimen Transfer of Flaw Size from Component to Simple Specimen

dhat_mean	**0.044**	flaw size units

Signal Variability Transfer Function

Only if a difference in the the signal variability is detected
Adjust the average signal due to additional variability in the Component compared to the Simple Specimen

Convert the signal variaiblity to flaw size variability using estimated slope coefficients

Fabricated Flaw in Component	0.0134	flaw size units
Fabricated Flaw in Simple	0.0009	flaw size units
Δdhat_var	**0.012**	Adjustment to Flaw Size Units (Difference)

Combining the Average Signal and Variaibility Specimen Type Transfers

dhat = dhat_mean - Δdhat_var

dhat	**0.032**	Simulated-Induced flaw size in Simple specimen for POD Demo

Appendix F. Inverse Case: Excel Spreadsheet Evaluating Similarity and Developing Transfer Functions

INVERSE CASE - Specimen Type Similarity and Transfer Function Calculations Input Output

Reference NASA TM "Guidebook for Assessing Similarity and Implementing Empirical Transfer Functions for Probability of Detection (POD) Demonstrations for Single-Hit Nondestructive Evaluation (NDE) Methods"

Scenario: Given a a96/95 flaw size derived from a previous POD study using simulated-induced flaws in a simple specimen, Estimate the representative naturally occurring flaw size in the component of interest

Summary of Analysis: Input a90/95 Flaw Size, and the Resultant Estimated Naturally Occurring Detectable Flaw in Component

a90/95 Flaw Size	0.040	From POD study of Simulated-Induced Flaws in Simple specimen (e.g., fatigue cracks in flat plate)
Component Flaw Size	0.072	Estimated Representative Natually Occuring Detectable Flaw Size in Component

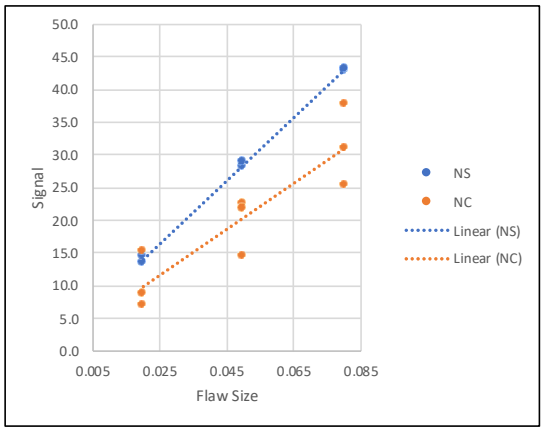
Transfer Evaluation Measurements used to Assess Similatiry and Specimen Transfer Function(s)

Simulated-Fabricated Flaws in Simple

Flaw Size (x)	Signal (y)
0.020	13.5
0.020	14.4
0.050	28.2
0.050	29.0
0.080	42.8
0.080	43.2

Simulated-Fabricated Flaws in Component

Flaw Size (x)	Signal (y)
0.020	8.8
0.020	15.1
0.020	7.0
0.050	22.5
0.050	21.7
0.050	14.5
0.080	31.1
0.080	37.8
0.080	25.4



Modeling of Transfer Evaluation Inspection Measurements

Estimate the Coefficients of Linear Models

Signal = ! 0 + ! 1*Flaw + "

Model Coefficients for Simple		**Model Coefficients for Component**	
! 1_hat_NS	484	! 1_hat_NC	352
! 0_hat_NS	4.31	! 0_hat_NC	2.82
s_NS	0.45	s_NC	4.71

Testing Similarity of Average Signal

Can we reliably detect a difference in the average signal for the same size Flaw in the Simple specimen compared to the Component?
If the confidence intervals do not overlap at the target flaw size, then we claim to detect an effect on the average signal response
ASSUME: Signal for Component is less than Signal for Simple specimen at Target Flaw Size

95% Confidence Interval of the Fitted Line at Target Flaw Size

Confidence Interval for Simple			**Confidence Interval for Component**		
y_hat_NS	23.7	at Target Flaw Size	y_hat_NC	16.9	at Target Flaw Size
t-value for NS	2.78	95% CI	t-value for NC	2.36	95% CI
delta_y_CI_NS	0.56		delta_y_CI_NC	4.29	
Upper CI_NS	24.2		Upper CI_NC	21.2	
Lower CI_NS	23.1		Lower CI_NC	12.6	

Is the Lower CI in Simple > Upper CI in Component? TRUE

If "TRUE" then we can detect a difference in the signal level at the target flaw size

Testing Similarity in Signal Variability

Can we reliably detect a difference in the signal variability between the Simple and Component specimens?
Based on the squared ratio of the RMSE from the regression models

(s_NC^2)/(s_NS^2)	107.6	
Cut-off value to Exceed	6.1	Finv, 95% one-sided test of Comp_var > Simple_var

Is the ratio of the squared RMSEs greater than the Cut-off Value? TRUE

If "TRUE" then we can detect a difference in the signal variability between the Simple and Component

Average Signal Transfer Function

Only if a difference in the average signal is detected
Adjust the average signal from the Simple to Component Specimen

Specimen Transfer of Flaw Size from Simple to Component Specimen

chat_mean	**0.059**	flaw size units

Signal Variability Transfer Function

Only if a difference in the the signal variability is detected
Adjust the average signal due to additional variability in the Component compared to the Simple Specimen

Convert the signal variaiblity to flaw size variability using estimated slope coefficients

Fabricated Flaw in Component	0.0134	flaw size units
Fabricated Flaw in Simple	0.0009	flaw size units
Δchat_var	**0.012**	Adjustment to Flaw Size Units (Difference)

Combining the Average Signal and Variaibility Specimen Type Transfers

chat = chat_mean - Δchat_var

chat	**0.072**	Estimated Equivalent Natually Occuring Detectable Flaw Size in Component

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
03/02/2022	Technical Memorandum	

4. TITLE AND SUBTITLE
Guidebook for Assessing Similarity and Implementing Empirical Transfer Functions for Probability of Detection (POD) Demonstrations for Signal Based Nondestructive Evaluation (NDE) Methods

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Koshti, Ajay; Parker, Peter A.; Forsyth, David S.; Suits, Michael W.; Walker, James L.; Prosser, William H.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
869021.01.23.01.01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER
NESC-TI-21-01657

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA/TM-20220003648

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified - Unlimited
Subject Category Mechanical Engineering
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This document provides guidance for assessing similarity relative to nondestructive evaluation (NDE) detectability based on Probability of Detection demonstrations when comparing different NDE inspection situations. Differences in inspection situations addressed by the similarity assessment methodology herein can include differences in the component being inspected, and differences in the conditions for the inspection. Additional differences in the inspection situation relative to differences in the method or procedure are not addressed in this document but can be assessed using NDE calibration or instrument standardization specimens.

15. SUBJECT TERMS
Probability of Detection; Nondestructive Evaluation; Handbook; Standard

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	61	(443) 757-5802